Enhancing Volunteered Geographical Information (VGI) Visualization with Open Source Web-Based Software

by

Vivien Deparday

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Geography

Waterloo, Ontario, Canada, 2010

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Recent advances in information technology such as Web mapping and location-aware mobile devices have enabled non-experts to create, use and share volumes of spatial data in an increasingly accessible fashion. Such user-generated spatial data is usually referred to as Volunteered Geographic Information (VGI). Two of the fundamental challenges associated with the exploitation of VGI relate to information overload and extraction of meaning. In order to deal with these challenges and improve the utility of VGI, this thesis investigates the potential of several interactive geovisualization techniques including filtering, dynamic spatial aggregation, linking and brushing, and tag-based visualizations. As a preliminary work to explore and structure the new research field of VGI, a framework of the different types of VGI is elaborated and followed by a review of the challenges and current solutions related to the utilization of VGI. Based on this review, a web-based prototype is developed to serve as a platform for the evaluation of selected geovisualization techniques. The prototype is then used in a series of workshops with rich citizen-generated data related to place-based community assets. The results of the case study show that the implemented geovisualization techniques enable users to find relevant subsets of information and to gain new insights on the data. Based on the potential shown by these results, future research directions are suggested.

Acknowledgements

I would like to thank all my colleagues, friends and family who supported me through this long but rewarding experience from Canada, France and other parts of the world.

I am particularly grateful to Dr. Rob Feick, my academic supervisor for not only his steady support and guidance throughout my academic work, but also for introducing me to pertinent and exciting topics of geographic information sciences. I would also like to thank him for his kindness that made this work and my integration in Canada even more enjoyable.

I would also like to thank Dr. Brent Hall for his judicious advice on my thesis, as my committee member, and for his help in setting up the initial field work with MapChat in the Bulkley Valley. I am also thankful to my readers, Dr. Peter Deadman and Dr. Doug Dudycha, for their insightful feedback at my thesis defence. I would also like to acknowledge Mike Leahy for his technical help with MapChat and for his support to make me adopt Linux completely.

The field work of this thesis would have not been possible without the participation and enthusiasm of the residents of the Bulkley Valley. I wish to express my gratitude to Dr. Ray Chipeniuk and to Morgan Hite, Anne Hetherington, Jay Gilden, and all other members of the Bulkley Valley Stewardship Coalition for their warm welcome in the Bulkley Valley and their help in setting up my field work. I wish to thank the study participants from the University of Waterloo and from the Bulkley Valley who gave their time and energy to attend my workshop and help my research.

I gratefully acknowledge the Canadian Network of Centres of Excellence for funding this research through the GEOIDE Project HSS-DSD17.

I would like to thank my family, my parents, my sister, brother-in-law and nephew for their essential support and presence despite the geographical distance. I am much grateful to my partner Drea for her unwavering support and patience despite my repeated excuses that kept me unavailable for long days and nights.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Problem Statement

Geographic information production has traditionally involved long processes that include data acquisition, processing, and dissemination phases that rely on expensive technology and expert staff. As a result, geographic information has been historically produced by either large public organizations such as the US Geological Survey (USGS) in the United States, l'Institut Géographique National (IGN) in France, and Natural Resources Canada (NRCAN) in Canada or by commercial data providers such as TeleAtlas and Navteq (Cowen, 2007).

This approach to spatial data production has important advantages. Organizations can capitalize on their expertise and equipment to realise scale economies and ensure uniform quality control regarding the completeness and accuracy of the data (Flanagin and Metzger, 2008; Goodchild, 2008). However, data production through a limited number of large organisations also presents some limitations. Profit potential or cost recovery often guides the data production process in the private and public sectors respectively. In the past, these factors often had the effect of restricting spatial data use to those who could afford to purchase a costly license or to produce it themselves. Lately, more public geospatial data are becoming available for free from sources such as USGS, NRCAN and the city of Vancouver as part of a larger movement to open public data (The Economist, 2010). However, this is not a global process yet and the data quality or currency may be inferior to commercial data, which can limit its potential uses. Moreover, the profit motive causes distortion in the supply of available data since data producers tend to favour the most profitable data such as popular areas, data themes with multiple applications or data themes that do not change rapidly (Goodchild, 2008). Similarly, because of the top-down approach of geographic data production, the data available typically lack local perspectives and are restricted to a single world view.

Such distortions in available data combined with the unequal access to GIS resources due to skills and funding barriers have important societal impacts. These have been debated in the "critical GIS" or "society and GIS" literature (Craig, Harris, and Weiner, 2002a; Pickles, 1995; Weiner and Harris, 2007). To tackle these inequitable impacts of GIS, some academics and practitioners have investigated new research avenues, methods and practices that were eventually identified as a new research area called Public Participation GIS (PPGIS). The PPGIS field encompasses a wide array of methods and projects ranging from mapping indigenous knowledge in developing countries to the definition of conservation areas in developed countries. However, common goals are the reduction of the unequal access to geographic data and software, and the collection and integration of local knowledge with official data in order to achieve a more democratic nature GIS. Until recently, such initiatives encountered mixed success due to technical and institutional barriers such as the high complexity and cost of GIS software (Elwood and Ghose, 2001; Esnard, 2007).

The advent of the World Wide Web (hereafter referred to as the 'Web') and especially Web mapping technologies, symbolized by the ubiquitous Google Maps, has made digital mapping accessible to novice users (Turner, 2006). Beyond the simple navigation of maps over the Internet, the recent Web technologies and practices, termed Web 2.0, allows users to create their own geographic data, known as "user-generated data" or "volunteered geographic information (VGI)" (Goodchild, 2007) that can then be easily shared with others. This phenomenon has significant impacts on geospatial data production and sharing as individuals can now act as data producers and users (Elwood, 2008a).

These technological and practical changes have an important potential to support the ideals of PPGIS and to help produce rich local and citizen-generated data that can complement official public and commercial datasets. However, the richness and complexity of citizen-generated spatial data makes their use and integration with official data challenging. For instance, the quality and credibility of VGI are difficult to assess as the data are produced by multiple amateur users in a distributed fashion. Moreover, due to its multi-authored nature, VGI can contain redundant and heterogeneous data. For instance, contributors may input similar data multiple times or they may use different terminologies within a same dataset. Furthermore, VGI often contains qualitative and subjective aspects inherent to human perception and description of space for which regular GIS models are not suited (Carver, 2003; Sieber, 2006). These challenges can severely limit the usability and utility of VGI for citizens as well as for its integration in decision-making processes and other official contexts.

Visualization and analytic tools have great potential to explore complex and multifaceted VGI without altering or simplifying its rich content. However, many of the tools that are available currently are designed to assist expert users to explore scientific data in a desktop computing environment through statistical graphing capabilities for univariate and multivariate data. As such, they have significant limitations in the VGI context where the nature of the data is different and the users are novices who operate over the Internet. Therefore, visualization tools need to be adapted or even re-invented to support novice and average users operating on the Web and its technical constraints. Indeed, only a few Web-based applications provide visual analytic tools for average users, as noted by Gorman (2006). Moreover, recent online mapping advances have not yet been able to deal with most types of VGI that contain qualitative and vague data and little work has been pursued to deal with the VGI specificities. Some early features can be found on the Internet such as the clustering of overlapping data. Nevertheless, experimentation and tests with end-users are required to assess the actual usability and value of such features and eventually develop new and improved methods to exploit the richness of VGI.

1.2 Research Objectives

By providing local and multiple perspectives, VGI has the potential to provide new insights on many topics and issues of importance to community members. On the other hand, this richness and particularities raise numerous challenges that limit its potential use. Hence, the overall goal of this thesis is to investigate methods to facilitate the exploitation of VGI and therefore improve its usefulness for citizens and possibly also planners and politicians. To do so, it is necessary first to identify the different types of VGI, the unique characteristics that contribute to its richness and also make its use challenging. This constitutes the foundation for subsequent development of methods and techniques that can improve the usefulness of VGI. Out of the numerous potential solutions, this thesis focuses on the use of fundamental visualization techniques to facilitate the browsing and exploration of VGI. Several techniques are selected and implemented in a Web-based prototype that aims to be usable by novice users. This prototype is then used as a test bed in a series of workshops where participants experiment and evaluate its features. Data on the usage of the features and participant feedback are collected in order to determine efficient techniques and areas where potential improvement and further research could be achieved. This thesis addresses the following research objectives:

- Review the literature on the Geospatial Web, PPGIS, and VGI to establish a VGI typology that is used as a foundation for the thesis.
- Review the challenges related to VGI use and visualization and their potential technical solutions.

- Design and develop a prototype Web tool that implements selected techniques to improve visualization of VGI.
- Test the prototype tool with a cross-section of potential users to assess the relative merits of these techniques in terms of usability and utility.

1.3 Study Area / Case Study

The Bulkley Valley, located in north western British Columbia, consists of an agricultural plain surrounded by mountains, rivers and forests. This setting offers a multitude of amenities such as scenic landscapes and outdoor recreational opportunities as well as urban facilities which attract an increasing number of residents and visitors. It also has an important potential for resource extraction industries such as forestry and mining. Due to these different factors, the Bulkley Valley is undergoing strong development pressure. Thus, some citizens have expressed concerns that the current and potential developments could negatively impact and possibly irreversibly damage important local amenities.

To address these risks, several groups of engaged local residents have attempted to take a proactive approach to the planning process. With the assistance university-based researchers, they have been involved in a participatory approach which engages volunteer citizens in the inventory of community assets. The inventory was performed by citizens using a Web mapping software tool called MapChat that allows users to annotate a common map to denote community assets such as trails and sensitive habitats and add descriptive dialogue relating to the assets' importance.

A series of workshops in the Bulkley Valley resulted in the collection of citizen-generated spatial data. These data are an example of VGI consisting of descriptions and opinions linked to map drawings that were contributed by volunteer citizens possessing very different skills and viewpoints. Local citizen groups were then eager to take the initiative to the next step and use the data collected to inform the planning process. However, they encountered difficulties browsing and using the data they collected that are symptomatic of the VGI challenges. Therefore, the data collected in the Bulkley Valley and the issues encountered by the citizens constitute a relevant setting for a case study. The data were used to feed a prototype Web-GIS tool that implements several visualization features. These features were evaluated by a cross-section of users in order to gather feedback on their usage and their subjective appreciation of the features implemented in the prototype.

1.4 Organisation of the Thesis

To achieve the research objectives presented in Section 1.2, the thesis is structured as follows. Chapter 2 first reviews the concepts of PPGIS, Web 2.0, neogeography and their linkages. Together, they constitute the origins of the Volunteered Geographic Information (VGI) phenomenon and its different definitions. To explore this phenomenon in more depth, a VGI framework that delineates the main types of VGI is developed and illustrated with a sample of current VGI applications. To restrain the thesis scope, one type of VGI and its related challenges are focused on. Some challenges are selected to be addressed in a case study. Chapter 3 describes the features developed and integrated in the software prototype MapChat Viz in order to explore possible solutions to the selected challenges. Chapter 4 presents the design and the execution of the workshops organised for the case study. Chapter 5 provides an analysis of the data collected during the workshops. Finally, Chapter 6 presents and discusses the findings and the limitations of the study in order to give recommendations to future studies.

Chapter 2
Literature Review and Research Framework

This chapter describes and links the different concepts and research questions underlying the thesis. Sections 2.1 and 2.2 provide an overview of Public Participation GIS (PPGIS), Web 2.0 and the neogeography phenomenon. These sections present some of their similarities and highlight how they both lead to the creation of spatial user-generated content. Section 2.3 explores several definitions of VGI followed with the development of a framework for examining VGI that is illustrated with a sample of VGI applications. The emerging research questions related to the nascent research field of VGI are reviewed in Section 2.4 with an emphasis on the questions addressed in this study.

2.1 Public Participation GIS

Consideration of the social, ethical, political and societal aspects of Geographic Information Systems (GIS) led to a vigorous debate among researchers in the 1990s (Craig, Harris, and Weiner, 2002b; Pickles, 1995). One of the main concerns was the elitist and exclusionary aspect of GIS technology (Pickles, 1999). Indeed, there are important inequalities in the access and use of spatial data and technology across society. GIS software, hardware and spatial data are expensive and require expert skills to be used efficiently which prevent numerous people and organisations from using them. Furthermore, the technological evolution of GIS has historically been directed most in accordance to the needs of customers with capacity to fund software development, such as large corporations, military and other large institutions. As a result, GIS were seen to reflect the views, values and interests of dominant sectors of the society and further marginalize minorities.

An important point of controversy between GIS critics (Lake, 1993; Taylor, 1990) and GIS proponents (Openshaw, 1991) is the fact that GIS embed positivist assumptions by using data that represent simple facts instead of more complex and multifaceted knowledge. In the critics' opinion, GIS show an overly simplified view of the world consisting of a single perspective that neglects multiple nuanced geographical realities and serves to "reduce complex societal processes to points, lines, areas, and attributes" (Sieber, 2006, p. 491). Moreover, other forms of information that can have important geographic dimensions such as history, emotions, and sacredness are hard to incorporate in GIS data models which tend to focus on the numerically tractable aspects of issues.

Besides the influence of the type of data used in GIS, the sources of data also have important consequences. There are a limited number of 'official' sources of data and these are mainly representative of the dominant and general views. Some communities whose values and interests differ may be under or not represented at all in the available data (Pickles, 1999). Without relevant data to advocate their case, some communities can be marginalized. Community perspectives and local knowledge are often neglected in official datasets either because they are simplified to fit into the data models, they do not match with the mandates of the organizations producing the data, or they are not financially lucrative. In this context, local knowledge or indigenous knowledge can be defined as 'value-based and traditionally intangible information' (Sieber, 2006), as 'knowledge that is unique to a given culture or society' (Warren, 1991), or finally as 'unique, traditional, local knowledge existing within and developed around the specific conditions of women and men indigenous to a particular geographic area' (Grenier, 1998). Their integration into data models is particularly important to move toward a more democratic use of GIS, as is discussed later in this section.

Overall, the assumptions and values embedded in available tools and data raise questions about the claimed objectivity and value-neutral nature of GIS. These issues along with unequal access to technology and data constitute the base of the critiques that considered GIS as an anti-democratic technology which reinforces existing divisions in society supporting a top-down, elitist approach to decision making in which most citizens cannot partake (Pickles, 1999). However, critics had to recognize that GIS was becoming more widely used in society. Therefore, to go beyond the simple criticism and build a constructive debate that would move the GIS field forward, academics and practitioners engaged in a research process that involved new methods, practices, and tools to address the critiques.

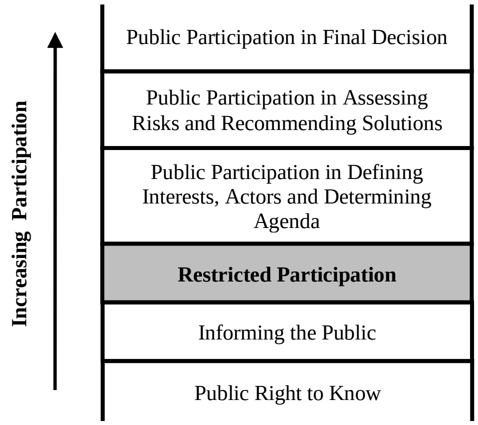
This new area of research has been referred to variously as GIS-2, critical GIS, Community-integrated GIS (CiGIS), participatory GIS (PGIS), or public participation GIS (PPGIS). Although these terms represent slight variations in the concepts (Weiner and Harris, 2007), the PPGIS term is now most widely accepted and will be used in the reminder of this thesis. The PPGIS approach is based on several core principles. First, there has been a concerted effort to disseminate GIS technology to marginalised groups such as grassroots and community-based organizations with the aim to empower them through enhanced opportunities ranging from the simply discovery of their environment to participation in formal decision making process. To reach this goal, GIS tools and practices needed to be adapted to ensure easier access to the technology for groups with different skills and resources (financial and otherwise). Furthermore, the integration of their local knowledge is critical to democratize GIS within these communities (Carver, 2003; Sieber, 2006).

PPGIS concepts have been applied in a wide range of contexts including neighbourhood revitalization (Elwood, 2002; Ghose, 2001), conflict management and collaboration among stakeholders (Balram and Dragicevic, 2006; Kyem, 2004; Weiner and Harris, 2003), land use and planning (Bojórquez-Tapia, Diaz-Mondragón, and Ezcurra, 2001), environmental management (Evans, Kingston, and Carver, 2004; Jankowski and Nyerges, 2001), indigenous territory claims (Dana, 2007), and numerous others can be found in (Craig et al., 2002a). However, it is important to note that the concept of participation does not refer to a single homogeneous method. To understand who benefits from access to GIS and why, and what are the appropriate methods, understanding the participatory process itself is essential. To do so, important questions need to be explored such as a) what is meant by participation?, and b) who is the public? (Dunn, 2007; Schlossberg and Shuford, 2005)

Participation or public participation is broadly defined as the process of involving citizens in various political, economic, social or other projects. A comprehensive review of the different ways to characterize the participation process is provided by Schlossberg and Shuford (2005). A common approach is to use the varying degrees of public involvement as characterized first by Arnstein (1969) as a ladder has been adapted by several authors including Carver (2003), as shown in Figure 2.1. The public participation in the decision making process can vary from a passive role with little participation at the bottom rung of the ladder to the role of decision makers who hold the power in the top rung. Between the two extremes, the level of participation and responsibility increases with upward movement. As noted by Carver (2003), participation is not always associated with empowerment. Unequal access to relevant information to advocate a case or barriers constructed by traditional power holders can hinder actual empowerment.

The public, also called PPGIS participants or stakeholders, generally comprises citizens who belong to one or several of the three following groups defined by Schlossberg and Shuford (2005, p. 18) as “those affected by a decision or program”, “those who can bring important knowledge or information to a decision or program”, and “those who have power to influence and/or affect implementation of a decision or program”. For example, stakeholders can range from municipal or regional government officials, developers, and scientific groups, to simple neighbourhood residents.

In some cases, neighbourhood residents may be represented by non-governmental organisations (NGO) or community-based organisation (CBO). Moreover, participants can have varying skills, interests, ages and cultural backgrounds.



Source: adapted from Carver (2003)

Figure 2.1: The public participation ladder

Therefore, to improve the chance of success and the impacts of a PPGIS initiative, it is essential to determine the type of participation and the members of the public that are relevant to its specific goal, context and its other social and cultural factors. For instance, identifying the relevant public is a critical task where finding a balance between a wide representation of the public with many participants and a smaller representation with fewer people but a more in depth participation can be difficult. Furthermore, even when relevant types of public and participation have been identified, numerous methods can be used. For instance in the spatial decision domain, methods can be facilitated using paper maps, or physical models to computer systems operated by a facilitator or self-operated. A comprehensive list of participation methods is presented in Rowe and Frewer (2005) and is characterized by a typology in Aysegul and Roche (2007). In the frame of this study, the focus is made on one of the PPGIS aspects namely the integration of local knowledge.

As described above, the data commonly used in GIS often do not incorporate local knowledge (LK). To take local perspectives into account, more efforts are being made to collect, use and integrate LK in GIS to complement “official” data (Dunn, 2007; Hall, Moore, Knight, and Hankey, 2009; Sieber, 2006). In this context, the term official refers to data that have been authored by professionals in government agencies or commercial mapping companies. The integration of LK can also help to improve the accuracy and completeness of official data since local people usually know their area better than other and may be able to provide valuable insights (Carver, 2003). Although necessary and powerful, the integration of LK with official data in GIS raises numerous challenges. Indeed, GIS are designed to handle and represent simplified factual information but are not suited for data that contain multiple perspectives and vagueness. Adapting GIS to incorporate such data or rearranging the data to fit GIS models are challenging issues (Sieber, 2006). Some LK may even be impossible to integrate and need to be taken into account by other means. The identification of these challenges and the design of potential solutions are major PPGIS research questions that are explored further in Section 2.4.

Through different PPGIS initiatives, various types of LK have been collected (as illustrated in the numerous case studies in Craig et al. (2002b)). However, actual empowerment through the use of LK has been generally limited (Weiner and Harris, 2007) due to technical, institutional, and political factors. The potential to collect and effectively use LK has been revolutionized by the advent of the Internet that enables individuals and groups to participate from anywhere, often at times that are convenient, and potentially in anonymous ways (Carver, 2003; Sieber, 2006). Moreover, recent Web-based tools are generally easier to use and can be operated directly by users with relatively little skill which enables more people to collect and share LK. The increasing use of Web-based methods in PPGIS is concurrent with major technology-driven changes in GIS practice, as discussed in the next section.

2.2 Web 2.0, Neogeography and their Synergy with PPGIS

In parallel to the practice-driven efforts in the PPGIS field to make GIS more accessible and suitable for public participation, a major technology-driven change is driving society in the same direction, namely the Internet network and the latest technology of the Web 2.0. The first broad public use of the World Wide Web was the publication of static Web pages that could be consulted by anyone connected to the Internet. The technology has recently evolved with the development of new design and programming methods known as asynchronous JavaScript and XML (AJAX). AJAX practices have allowed the creation of an ever more interactive and dynamic Internet where a portion of Web page can be changed or refreshed without reloading the entire page, thereby allowing users to interact dynamically with the display and with other users (Mahemoff, 2006).

Beyond the technological improvement, these evolutions have favoured the creation of new practices, design patterns and business models, grouped under the generic term Web 2.0 (see Table

2.1). A number of the Web 2.0 core concepts are particularly relevant in the GIS and PPGIS fields. In the Web 2.0 model, the Web is used as a platform (O'Reilly, 2005) where technologies and increasing bandwidth permit the development of software as a service. These “software services” are directly available on the Internet and usable through a simple Web browser. This dramatically reduces the need for users to purchase expensive hardware and software and as a consequence it broadens access to technology. Web 2.0 concepts also emphasize the use of an “architecture of participation” to harness the collective intelligence (Tapscott and Williams, 2006). In the Web 1.0 realm, users were only able to consult information on the Internet. It was a one-way information flow from the sponsor or facilitator (e.g. municipalities, NGO, etc...) to the public as defined by Rowe and Frewer (2005). Thus, only the types of participation corresponding to the first two rungs of Figure 2.1 are possible on the Internet. Enhanced Web 1.0 applications enable the possibility to survey the public by enabling a flow of information from the public to the facilitator. However no real interaction was possible. Most recent Web 2.0 applications enable full interaction between the facilitator and the public and even in between the members of the public themselves (Hall and Leahy, 2010).

Web 1.0	**Web 2.0**
Static	Dynamic
Publishing	Participation
Producer-centric	User-centric
Centralized	Distributed
Close-coupling	Loose-coupling
Basic	Rich

Source: adapted from Maguire (2007)

Table 2.1: Some differences between the Web 1.0 and Web 2.0

Architectures of participation can serve profit making contexts such as commercial sites like Amazon (www.amazon.com), but they also have the potential to facilitate the higher rungs of the participation ladder. They can also be used to get the best and most suited minds to solve issues such as finding a disease cure or elaborating a climate change model. For instance, the Website Innocentive (www.innocentive.com) is designed to assist contacts between solutions seekers and problem solvers for complex scientific issues. The same idea underlies the involvement of local people for the collection of LK. As mentioned in Section 2.1, local people can be the most suited to contribute LK (Carver, 2003). Another important aspect of the Web 2.0 model is the fact that applications are built of “a network of cooperating data services” (O'Reilly, 2005). Therefore, a

Website can directly use multiple remote data services and does not need to replicate them. For instance, a Website can retrieve the daily weather forecast from a remote Web service and integrate it seamlessly in its interface.

Such technologies and concepts are widely used by geospatial technology (Peng and Tsou, 2003). One of the main applications is the creation of distributed network of geospatial data providers. Various sources of data are available over the Internet through remote services and can allow users to create maps based multiple remote data sources. These concepts of cooperation and distribution extend to other aspects of the GIS such as computing power (e.g. grid computing allows networked computers to share their power), and software development (e.g. open source GIS are collaboratively developed by many developers) (Sui, 2008). The distribution and cooperation between system and people can help address the issues of inequitable access to data and technology.

Web 2.0 concepts and technologies are beginning to have a dramatic influence on the use of GIS in society. Indeed, they led to the democratization of GIS tools and of map-making. The popularisation was initially started with the ubiquitous Google Maps and Yahoo! Maps and their application programming interfaces (APIs) that allow users to develop customized mapping tools (Gibson and Erle, 2006). These new tools and techniques combine the complex techniques of cartography and GIS and place them within reach of people without formal training in map-making or use of spatial information. This phenomenon is sometimes identified as neogeography (Haklay, Singleton, and Parker, 2008; Turner, 2006). As a result, the Internet is replete with maps on various topics created by amateur cartographers (White, 2008). Making GIS tools usable by novice users and therefore broadening the GIS user base directly supports PPGIS goals. Overall, the Internet and the Web 2.0 concepts and technology offer tremendous potential to enhance the impact of PPGIS initiatives from which few projects have started to benefit (Bugs, Granell, Fonts, Huerta, and Painho, 2010; Hall, Chipeniuk, Feick, Leahy, and Deparday, 2010; Rinner, Keßler, and Andrulis, 2008).

2.3 Volunteered Geographical Information

2.3.1 Definition

These new technologies and practices have led to the development of a number of tools that enable novice users to produce their own geographical data. For instance, casual users can upload their hiking routes from global positioning systems, they can geotag their photos, or pinpoint on a map place-based experiences and opinions. The term Volunteered Geographic Information (VGI) (Goodchild, 2007) is often used to refer to these data that have a spatial component and that are created by volunteers. Others terms include user-generated content (UGC) coined during the Web 2.0 phenomenon or collaboratively contributed geographic information (Bishr and Mantelas, 2008). In the term VGI, the word "volunteered" is sometimes discussed (Sieber, 2007; Tulloch, 2008) since the data collection may not be "volunteered" but rather facilitated or even collected unbeknown to the "volunteers" as further described in Section 2.3.3. However, VGI is to date the term the most widely adopted by academia. The term VGI is also used beyond the designation of a type of data to refer to the whole phenomenon of collaborative creation of spatial data, a new research field or a group of application as defined by Tulloch (2008, p. 161):

> "VGI applications as those in which people, either individually or collectively, voluntarily collect , organize and/or disseminate geographic information and data in such a manner that the information used by many others."

Therefore, in the context of this thesis the term VGI will be used to refer either to the type of data or the related phenomenon and research field.

The development of the VGI phenomenon has been driven by complex and intertwined factors. As presented in Sections 2.1 and 2.2, two main influences can be identified, namely a human-driven push coming from the PPGIS tradition and a technological pull coming from the Web 2.0 and neogeography. First, major technological breakthroughs such as the enhanced interactivity of the Internet and the democratization of Web-based geographic technology have enabled the VGI phenomenon. Secondly, the collection and integration of LK promoted by PPGIS research practices has constituted a push towards the emergence of VGI. The clear parallel between the PPGIS, GIS, and the VGI field has led academics to investigate the relationship between them (Elwood, 2008b, 2009; Tulloch, 2008). However, the identification of their differences and similarities is a difficult task due to their multiple intertwined conceptual and technological aspects. After early investigation, it appears that VGI and PPGIS do not correspond exactly (Elwood, 2008b; Tulloch, 2008), as VGI stresses applications and information whereas PPGIS emphasizes process and outcomes. Despite these differences, the synergy between the two fields offers opportunities to allow both fields to move forward. For instance, Elwood (2008b) outlines a number of research questions and findings from the PPGIS field that could inform VGI research, such as the social and political aspect embedded in VGI creation. Similarly, Elwood (2009) explores the GIScience aspects that are relevant to the VGI field. Elwood identifies three main linkages, the handling of heterogeneous data, qualitative data, and dynamic forms of data.

Once simply considered as the hobby of amateur cartographers, VGI has now acquired a real importance in the production and dissemination of geographic data. It provides an alternative and a complement to the regular sources of geographic data. Indeed, it "has greatly enriched our potential for characterising the specificities of localities that central agencies either lack the resources, mandate or interest to collect and publish" (Hall et al., 2010). Furthermore, the interconnection of these multiple sources of data through the use of Web 2.0 technologies constitutes a patchwork of knowledge that is quickly and broadly disseminated (Goodchild, 2007). Therefore, VGI has become a unique source of data that can provide researchers and citizens with new insights on the world and the society they live in.

To explore further the nature and the extent of the VGI phenomenon and give a more concrete understanding of the research area, it is relevant to review and characterize practical examples and applications (Tulloch, 2008). This is a necessary first step towards the elaboration of a VGI typology which is used to structure the reminder of this thesis. However, GIS tools needs to be adapted and re-thought to support the increasing volumes and varieties of VGI. Hence, the definition of VGI and its different types first need to be reviewed and investigated.

2.3.2 Development of a VGI Typology

As mentioned by Elwood (2008b, p. 177), "descriptive efforts to characterize VGI and its implications" are critical to structure future research on VGI. Indeed, VGI encompasses such a wide spectrum of types of projects and data that certain approaches may be relevant in some situations but not be appropriate in other circumstances. Thus, a framework is necessary to structure the investigation of the various challenges and their related solutions according to the different types of VGI. There are many ways of describing and characterizing the VGI phenomenon depending on the aspects the research focuses on. Since this thesis focuses on enhancing the visualization of VGI and improving its utility in general, the typology developed in the following pages focuses on the nature of the volunteered data and the purpose for which it has been contributed. These two aspects are intrinsically linked to the challenges raised in the exploitation of VGI and their potential approaches and solutions.

First, a set of parameters related to the nature of the data and its intended use was developed by reviewing previous literature on PPGIS, neogeography and VGI. These parameters were then used to review and characterize VGI applications. Developing a full catalog of VGI applications as recommended by Tulloch (2008) would be ideal. However, in the frame of this study, considering a set of examples selected across the VGI continuum was deemed to be an appropriate alternative. The review and characterization of this sample of VGI led to the identification of the main VGI types. Section 2.3.4 presents the main types illustrated with the most remarkable examples.

It is important to mention that this classification process is only one of the many possible ways of categorizing VGI. It does not attempt to and cannot capture the entire complexity of the VGI phenomenon. Its goal is rather to emphasize aspects that are critical for the study presented in this thesis and to provide a basic context to structure some of the future research around VGI. In order to develop the set of parameters and the typology, it is first essential to review existing frameworks and the nascent literature on VGI.

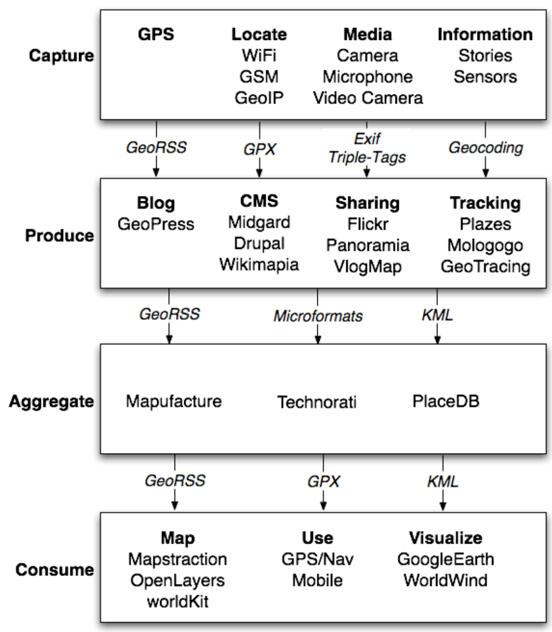
2.3.3 Elaboration of a Set of Parameters to Characterize VGI

In order to differentiate and characterize VGI, it is first necessary to elaborate a set of parameters that allows analysis of the similarities and differences across the VGI panel, and subsequently to identify the main types of VGI. The set of parameters is based on the review of the previous literature in the PPGIS, neogeography and VGI fields as well as on an analysis of the VGI landscape. The parameters are organized in two levels. First, high level parameters are used to describe the fundamental differences in the nature of the different types of VGI. This level is represented as a VGI continuum described in Section 2.3.3.2 and represented in Figure 2.4. The second level represents the secondary parameters that are used to characterize the types identified along the VGI continuum. They are described in Section 2.3.3.3 and inventoried in Table 2.2.

2.3.3.1 Review of Existing Frameworks

The literature on VGI provides various elements of description and characterisation of VGI. However, to date, there is no framework describing and characterizing the broad landscape of VGI and the only frameworks related to VGI are found in the PPGIS and neogeography literature. These frameworks do not focus on VGI but highlight various aspects that are important to its related fields. However, when reviewed with a perspective focused on the nature of the data and its intended use, they can bring essential blocks for the development of a VGI typology.

One approach coming from the neogeography perspective is elaborated by Turner (2006) and presented in Figure 2.2. This framework is based on the successive steps of data management that VGI goes through between its capture by volunteers and its consumption by an end-user. For each of these steps, different technological tools and related data formats are presented. This framework gives a good overview of some of the different technologies used in the production cycle of VGI in the context of Web 2.0. One drawback of the framework is that tools and technologies are evolving very rapidly so the framework would need to be updated regularly to present up-to-date information.



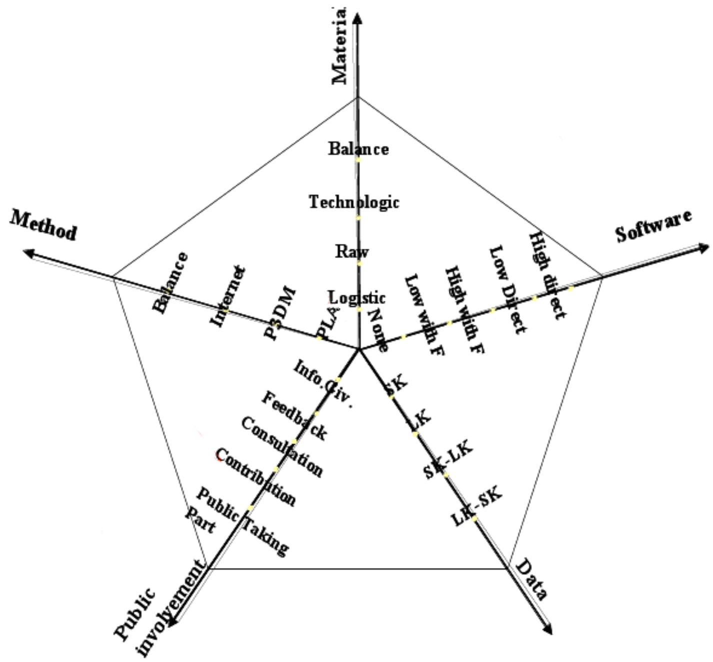
Source: Turner (2006)

Figure 2.2: Geostack

In the context of this thesis, it is important to note the different means of capturing geographic information as they directly influence the nature of the data such as its quality or format. On the other hand, since the focus is on neogeography tools and therefore highlights a highly technological type of VGI production, it leaves aside VGI that is created through simpler processes such as drawing on a paper or digital map, which also needs to be considered for this study. Therefore, in the typology developed here, both technological methods and simpler approaches are considered.

Other frameworks that are relevant in the study of VGI are those produced in the PPGIS literature. Numerous PPGIS frameworks examined the participation and especially the type of participation and nature of 'the public' as presented by Schlossberg and Shuford (2005). Recently, the Internet is often used to create the information through a usually flexible and anonymous process. It is therefore difficult to specifically identify 'the public' and approaches should assume a generally varied public with both novice and expert users. In terms of participation, three way communication (between the facilitator, the public and within the public as described in Section 2.2) is usually enabled by recent VGI applications. Therefore, these two aspects do not allow significant differentiations of VGI types and are not central to the objective of the typology elaborated here. A PPGIS typology was developed by Aysegul and Roche (2007) and is presented in Figure 2.3. It also includes considerations of the nature of the participation process and the public. They are characterized by several parameters, one for public involvement (see corresponding axis in Figure 2.3) similar to the ladder presented in Figure 2.1 and three parameters to characterize the technology employed ('Material' and 'Method' axes in Figure 2.3) and the interaction between the technology and the users ('Software' axis). As mentioned earlier, these elements do not allow significant characterization in the recent VGI context as the delivery method employed is mainly the Internet with a high interaction between the participants and the tools. However, the last parameter called "Data" is particularly relevant here as it characterises the data used in a study on a continuum anchored by scientific knowledge (SK) on one end and local knowledge (LK) on the other.

LK usually refers to data that are contributed by the public during participatory processes (see definition in Section 2.1). SK usually refers to official and scientific data (see definition in Section 2.1) created by a government or a private sector agency that serves as a base layer on top of which participants can contribute their LK. Thus, in this context, SK is not a type of VGI, as it is not contributed by volunteers. However, in the more recent context of VGI, some applications enable casual users to create SK. With this new perspective, the distinction between SK and LK can be transposed in the context of VGI. As such, it is used as a fundamental parameter in the development of the VGI typology as described in the next section.



Source: adapted from Aysegul and Roche (2007)

Figure 2.3: PPGIS typology

2.3.3.2 VGI Continuum

The typology elaborated for this thesis is based on a VGI continuum that covers the range of VGI types. This continuum is built on two levels. First, as noted by Elwood (2008b), there are two main purposes for the production of VGI. VGI can be contributed in order to complement or update official data often when official data are incorrect and/or not complete enough. In this manner, the interconnection of VGI and official data leads to the creation of a "patchwork" of knowledge as described by Goodchild (2007). Sometimes, beyond the simple completion of official data, VGI aims at replacing existing official data sets in order to liberate users from licensing and financial constraints. As such, these data are usually structured similarly to official data and conform to usual GIS models. The second main purpose behind VGI production is to "enable completely new forms of knowledge production, fostering new social and political practices" (Elwood, 2008b, p. 176). For instance, this practice can lead to the production of opinion-based, political or personal data that were almost nonexistent in regular GIS. These two types of VGI are radically different as one follows usual GIS knowledge structures, whereas the other one explores geography of a new kind. These two types are respectively termed “Conventional GIS knowledge” and “Unconventional spatial knowledge” in this study and are represented above the VGI continuum in Figure 2.4.

A second level of characterisation is based on the fundamental differences that exist in the nature of the contributed knowledge. The construction of this level is based on an adaptation of the SK and Local Knowledge LK distinction made by Aysegul and Roche (2007) as presented in Section 2.3.3.1. Within the VGI typology, SK refers to knowledge that is contributed by volunteers and that either is scientific by nature or describes the world in a quantitative or scientific manner. It is typically centred on measurable or quantifiable data such as soil classifications, population density, street networks, and so on. LK refers to knowledge that is related to a local perspective. It can consist of conventional GIS knowledge and/or of qualitative opinions or perceptions that are related to a particular point of view. Finally, a third type completes the full range of VGI namely the knowledge generated for personal purpose to share with friends, relatives or co-workers. This type is termed “personal knowledge” (PK). These three types of VGI are represented in a VGI continuum in Figure 2.4 and are further characterised in the next sections.

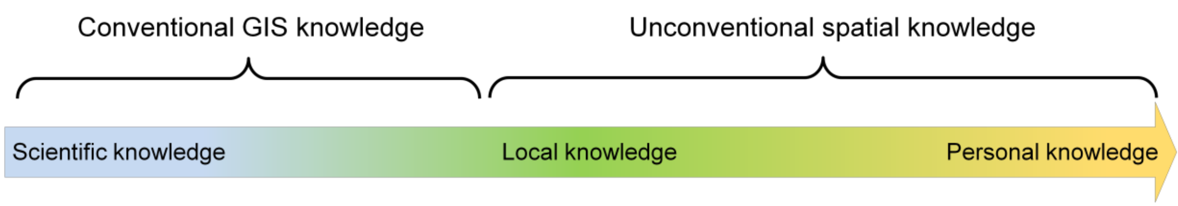


Figure 2.4: VGI continuum

2.3.3.3 VGI Characteristics

The VGI continuum presents the main types of VGI. These main VGI types are described and characterised according to the nature of the contributed data and the influence of the data contribution process on the nature of the data. This approach provides an important basis to structure future research as it allows the development of research questions that are relevant across the full continuum or only for one portion of it.

The traditional distinction between spatial and attribute data is used to structure the discussion that follows. However, it is important to note that even this distinction is challenged by the VGI phenomenon. Indeed, the spatialization or geoparsing of texts (see section 2.4.4) by the massive use of geocoding can transform any textual spatial reference into spatial information. Thus, written and spatial information are linked them together into an almost indivisible way as the spatial reference does not exist without the written one. Nevertheless, this distinction seems still relevant to organize the characteristics. An overview of the criteria structure is available in Table 2.2 and further developed in the reminder of this section.

First, the method of capture of spatial data is an essential characteristic as it is closely related to the accuracy of the generated data. Some of the most common tools amateurs use to create spatial data include:

- Global Positioning System. These devices have evolved from expensive professional devices to consumer devices that are available as stand alone or integrated into cars and phones. GPS are one of the main tools to generate geographic data even though their accuracy can vary widely.
- Location through the use of the Internet or cellular phone networks. The Internet network can provide approximate locations by comparing an address IP (Internet Protocol) against a database associating IPs and locations or also by using the triangulation of signals coming from wireless base stations. The mobile phone network can also permit triangulation methods. These methods are detailed in Turner (2006). These three methods and GPS are more and more combined together to locate mobile devices such as smart phones or personal digital assistants (PDA). Therefore, it can become hard to track back what has been exactly used to locate the device but the accuracy is still usually provided.
- Geocoding. Geocoding refers to the process of locating a place name or an address on a map. It is based on the use of gazetteers which associate place names and addresses to geographic locations. It has been democratized by the Google Maps and other publicly available interfaces that make the process almost transparent to end-users. As mentioned earlier, a technique related to geocoding is geoparsing (Scharl, 2007). It consists of scanning through written documents and extracting all the geographical references to locate them on a map by using geocoding techniques.
- Drawing on a computer-based map. This is a widespread method in Web mapping where a user can add new content to an overlay layer on a map that presents base data such as satellite imagery. The drawing can be limited to point feature in some applications, whereas other allow for complex lines and polygons. Sometimes, instead of being able to draw, users can simply select a feature already existing on a map and attach a comment to it.

- Drawing on paper maps. Despite the advent of new technologies, the use of paper maps to collect VGI remains essential in some contexts as it is significantly cheaper and more intuitive for non computer literate people (Al-Kodmany, 2001). As a testimony of the practicality of paper maps, the OpenStreetMap walking papers (http://walking-papers.org/) project allows users to print paper maps that they can draw on in the field and later scan then to generate automatically digital information.

- Twitter. Twitter is a recent Web 2.0 application. It was not designed for geographical applications but it can allow users to contribute geospatial information. Twitter is a micro-blogging platform that allows users to post message up to 140 characters to share their life or thought of the moment with others. This message can simply include a place name tag that will then be geocoded on a map. This method of information sharing has been used widely in some recent crisis events such as the Iranian elections, and the earthquakes in Haiti and Chile (Tesquet, 2009). Twitter is currently adding a functionality that will allow users to locate their post automatically through the location of their mobile devices.

A second important parameter of the spatial component of the data is the type of features supported by the applications. Some only allow points whereas others allow for lines and polygons. This aspect has important consequences on the richness of the data collected but also in the challenges to analyse and interpret it (see Section 2.4.5).

The second set of criteria relates to the nature of the data that are dedicated to the text component. Alphanumeric data generated through VGI capture methods tend to have a more flexible structure to leave enough freedom to the users to express their reality. It is essential to characterize this degree of freedom as it gives insights into the purpose behind the data collection and on the different methods for handling the data. First, to express flexibility in the data structure, the attribute data can be characterized as structured or unstructured. Unstructured data allows users to associate form free comments and opinions linked to spatial features. Structured data refer to attributes that conform to a range of values on nominal, ordinal, interval or ratios scales. VGI can also have a combination of structured and unstructured text associated with spatial features. For instance a Website that allows users to express reviews and opinions about restaurants can have free text for the review and a set of categories for the type of food. It is important to note that in recent applications data related to spatial features are not only in text format, but they can also be pictures and videos. In these cases, the differentiation structured/unstructured may not apply as readily.

Another characteristic of attribute data are their relative subjectivity or objectivity. As mentioned by Tulloch (2008), they can be purported facts which are objective data or offered opinions on an issue which is subjective. This criterion has to be considered carefully as some information could be presented as purported facts by one community but another community could have a different perspective on the same fact. Thus, the terminology objective/subjective will be used.

The second group of parameters is related to the characteristics of the process of the data contribution. The first characteristic revolves around the use of the word "volunteered" in the definition of VGI. As mentioned in Section 2.3.1, the contribution is not always entirely volunteered and different degrees of willingness can be determined. The first degree is when the contribution is actually freely volunteered and made available to everyone. A second degree of willingness can be identified when people are asked to participate, for instance for the collection of local knowledge within a planning process. Seeger (2008) coined the term of facilitated VGI (f-VGI) for this type of contribution. Another type of contribution is one intended for only certain recipients like a circle of friends (Elwood, 2008b; Sieber, 2007). This is common in social network applications. It is important to note that even though users may want to share this information with only a limited number of people, it is sometimes available for everyone either because of the software design or because of user errors in software manipulations. Finally, the last level is when contributors are mainly unaware aware that they are contributing (Elwood, 2008b; Tulloch, 2008). For instance when using GPS-enabled phones to take pictures and post them on the Internet, some users are not aware that the pictures location is embedded in the image files and can be used by ill-intentioned individuals. This constitutes only one example of what has now been termed cybercasing (Friedland and Sommer, 2010). It also echoes the earlier critique of GIS as a means of surveillance from the "GIS and Society" debate (Pickles, 1991). Obermeyer (2007) describes this phenomenon as geoslavery where people agree with the surveillance for the benefit it brings despite the important threat on their individual privacy.

The second characteristic inherent to the VGI collection process is the degree of interaction and involvement made possible to the users. As presented earlier, this concept is central in many PPGIS frameworks where it is represented by several versions of the public participation ladder (Schlossberg and Shuford, 2005). However, a simpler approach adapted from the Rowe and Frewer (2005) typology presented in Section 2.2 considers information flows in the context of this typology. One way flows from the facilitators to the users where users are only informed and do not contribute data is not considered here as it is not relevant in a VGI context. The second type of one way information flow, where users contribute information in a survey approach but do not see the result of their work, is one case of VGI. However, even though this type of interaction was quite common in early PPGIS initiatives, it is becoming rare in recent VGI applications. With two way information flows, users contribute data and are able to see the results of their work combined with the contribution of all other users. This interaction is now very common since producers and users are often synonymous in the VGI context (Goodchild, 2008).

There are actually now VGI applications where there is no determined facilitator but only a community of users. Thus, these two types of interaction can be complemented by a n way flow interaction that is not present in the Rowe and Frewer (2005) framework. A n way flow refers to applications where users can also cooperate in the data collection by commenting or editing each others' contribution. Therefore, three levels of interaction can be used for VGI, namely one way from the users to a facilitator, two way between a facilitator and users, and finally n way where communication expands to include user to user flow.

Data				Process	
Spatial		Attribute			
Capture	Feature	Structured/ Unstructured	Objective/ Subjective	Interaction	Nature
- Selecting - Drawing - GPS - Geocoding - Location through networks - Twitter	- Point - Line - Polygon	- Structured - Unstructured - Other formats	- Objective - Subjective	- 1 way - 2 way - n way	- Volunteered - f-VGI - Private - Unbeknownst

Table 2.2: Selected VGI characteristics and their possible value.

The next section uses these parameters to describe and characterize the main VGI types identified along the VGI continuum.

2.3.4 A VGI Typology

The review of a sample of VGI applications enabled the delineation of three generic VGI types of VGI. Each type is located along the VGI continuum and described by set characteristics. Slight variations in the characteristics within each type led to the identification of several subtypes. These types and subtypes are presented and described in the following section. The characterization of each VGI type is illustrated with a reference to popular current VGI applications. All the selected examples are positioned within the framework graphically represented in Figure 2.5.

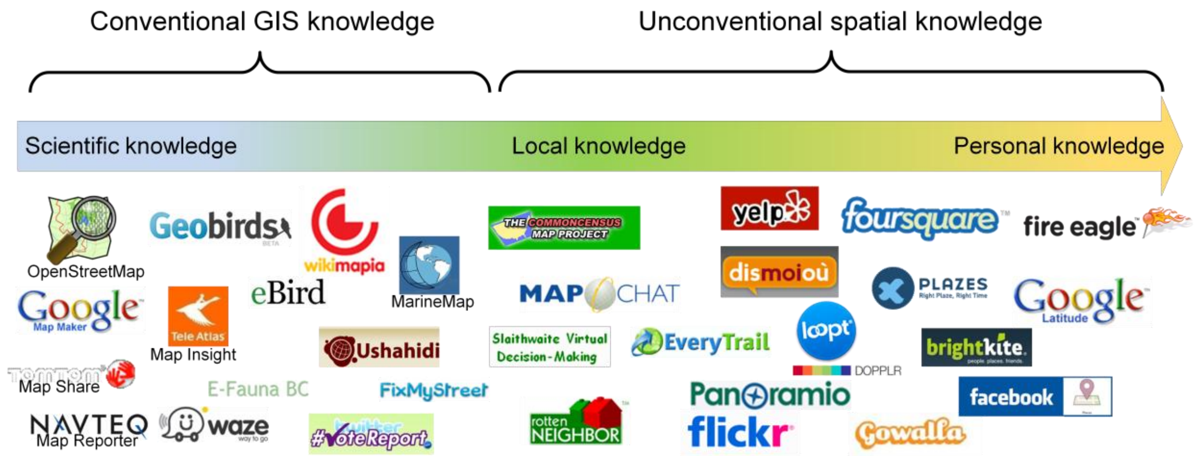


Figure 2.5: VGI framework

Scientific knowledge VGI Type 1	Local knowledge VGI Type 2	Personal knowledge VGI Type 3
- volunteered - Objective - Structured - 2 way or n way - Digitizing, GPS, twitter - Only points or points, lines, polygons	- facilitated-VGI - Subjective - Unstructured - 1 way, 2 way or n way - Selection, Drawing, geocoding - Points, lines, polygons	- Kept private - Subjective - Unstructured - n way - Location through networks, geocoding - Only points

Table 2.3: Table summarising the main characteristics of each VGI type.

2.3.4.1 Scientific Knowledge –Type 1 VGI

Description

The left end of the continuum is anchored by Type 1 VGI which, in many respects is similar to data that have been traditionally been created by professionals. It concerns data that are predominantly structured and objective (see Table 2.3). Type 1 VGI is contributed in a volunteered process by users ranging from amateurs to professionals. They can usually collaborate in a three way fashion, specifically they can contribute data, get and use the resulting datasets, and importantly they can correct others' contributions which, as a result, greatly improves the data quality. In some cases, this may also lead to deliberate error creation but VGI projects usually have safeguard to avoid that. To create the data, precise means of capture are usually favoured such as GPS receivers, however to allow more people to participate, simpler methods are sometimes used. Type 1 VGI includes base data layers such as street network, building footprints or point of interests (schools, airports, etc). The generation of such datasets through collaborative effort is usually motivated by the possibility to create open datasets (see OpenStreetMap (http://openstreemap.org) example below) as an alternative to restricted and expensive commercial ones. However, it is important to note that VGI use can sometimes be restricted by licences (see commercial map maker examples below). Another example of Type 1 VGI is research data such as climate change observations or bird sightings (see eBird example below). Such data particularly benefit from the VGI production approach as they are very costly, if not impossible, to collect without the help of numerous volunteers.

Examples

OpenStreetMap

OpenStreetMap (OSM) is a global initiative to create collaboratively a free-to-use world dataset using the efforts of participants from all over the world. The project was initially focused on streets but has now expanded to many other types of feature such as trails, bike lanes, parks, rivers, house numbers, restaurants and other point of interests. The volunteered collaborative process occurs in a Wikipedia-like method. Users can contribute new features to the map or edit and correct what has been done by other users. The usual process to create a new feature is first to collect GPS tracks in the field, upload them on the OSM server and finally edit them with either a Web client directly accessible from the OSM interface (Figure 2.6) or with more advanced standalone software like JOSM (Java OSM) editor. Another method to input data is the digitisation of features directly based on an aerial imagery. However, this can be performed only with data sources that have a license compatible with OSM. For instance, Yahoo! Imagery granted the OSM project the right to use its satellite imagery Web service. Similarly, several government mapping agencies such as the US Census as well as some commercial firms (e.g. AND (www.and.com)) have granted OSM access to key base map datasets. In contrast, any feature that is digitized or reproduced from non-public-domain data like Google Maps is prohibited by the OSM community. Once the spatial data are created, attributes are added to the data to describe the features. Although theoretically free, most users follow an attribute structure elaborated collaboratively by the OSM community.

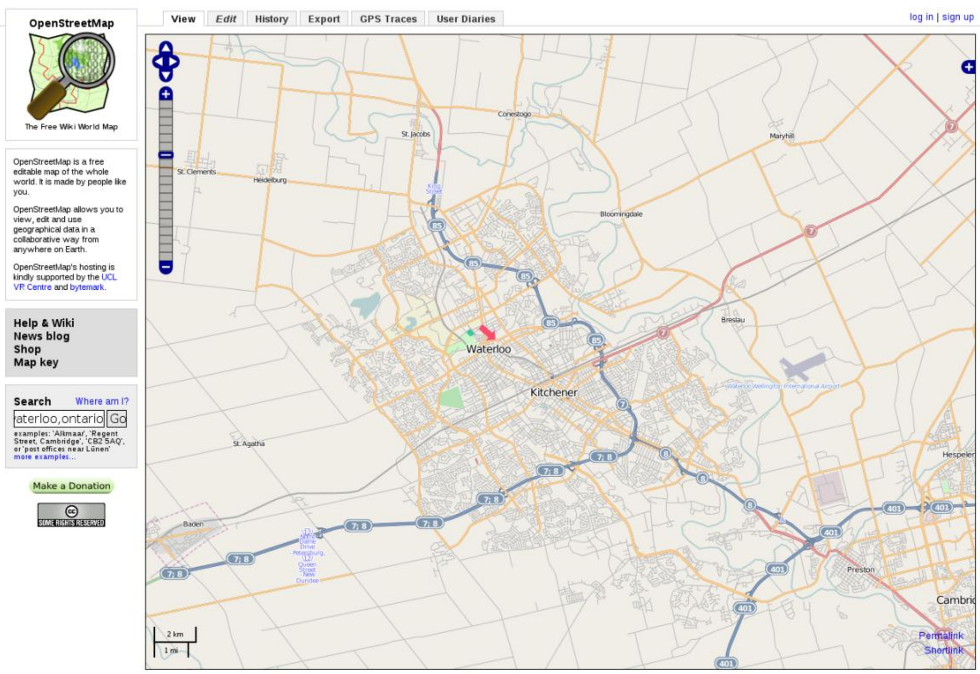
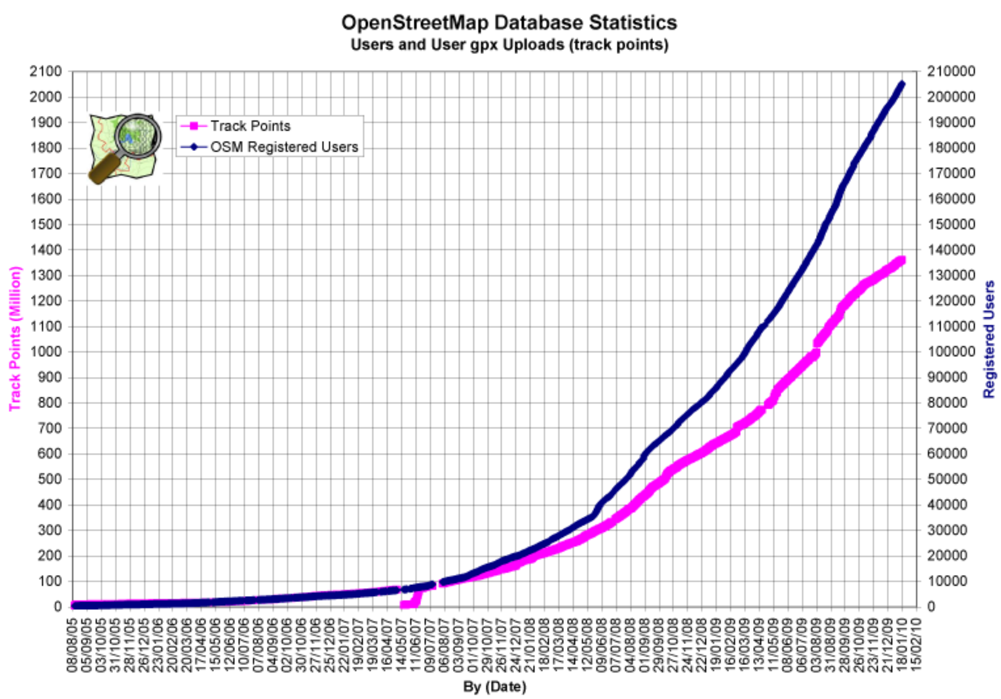


Figure 2.6: OpenStreetMap centred on the Kitchener-Waterloo area

The main incentive for the users (e.g. members of the public, corporations, non-profit organizations, researchers) to contribute data is that all the data contributed are going back to the community either as Web maps and can also be exported in various formats for personal uses under the Creative Commons Attribution-ShareAlike license (http://creativecommons.org/licenses/by-sa/2.0/). However this license has limitations when used with geographical data and new licenses such as the Open Database License (http://www.opendatacommons.org/licenses/odbl/) are under development to deal with these issues, as detailed in Haklay and Weber (2008) and in Section 2.4.7. Licensing issues are especially important and complex in the VGI context as the data are created by many users and are usually ultimately meant to be shared and reused. As it happened in the open source software world, the definition of standard licenses applicable in the context of VGI is necessary to increase the ease of use and consequently the utility and the visibility of VGI. OSM is supported by a growing dynamic community (see Figure 2.7) that organizes social events such as map parties (Perkins and Dodge, 2008) and even a yearly OSM conference called State Of The Map (http://www.stateofthemap.org/). As a result, an important amount of data has been collected and is rapidly increasing (see Figure 2.7) and a video made by ITO world (http://www.itoworld.com/)). This success has made OSM an emblematic VGI application which benefits from a lot of press coverage in specialist and mainstream media (http://wiki.openstreetmap.org/wiki/OpenStreetMap_in_the_media) that then attracts more new users. The creation of companies such as CloudMade and Geofabrik (http://cloudmade.com/ and http://www.geofabrik.de/) based on business around OSM asserts this success. This project illustrates the power of VGI. For the well mapped areas of the world, it offers an alternative to the cost-prohibitive commercial dataset (Haklay and Weber, 2008) and for areas of the world that are left aside because they are not considered profitable by commercial mapping companies, it sometimes offers the only widely available mapping.

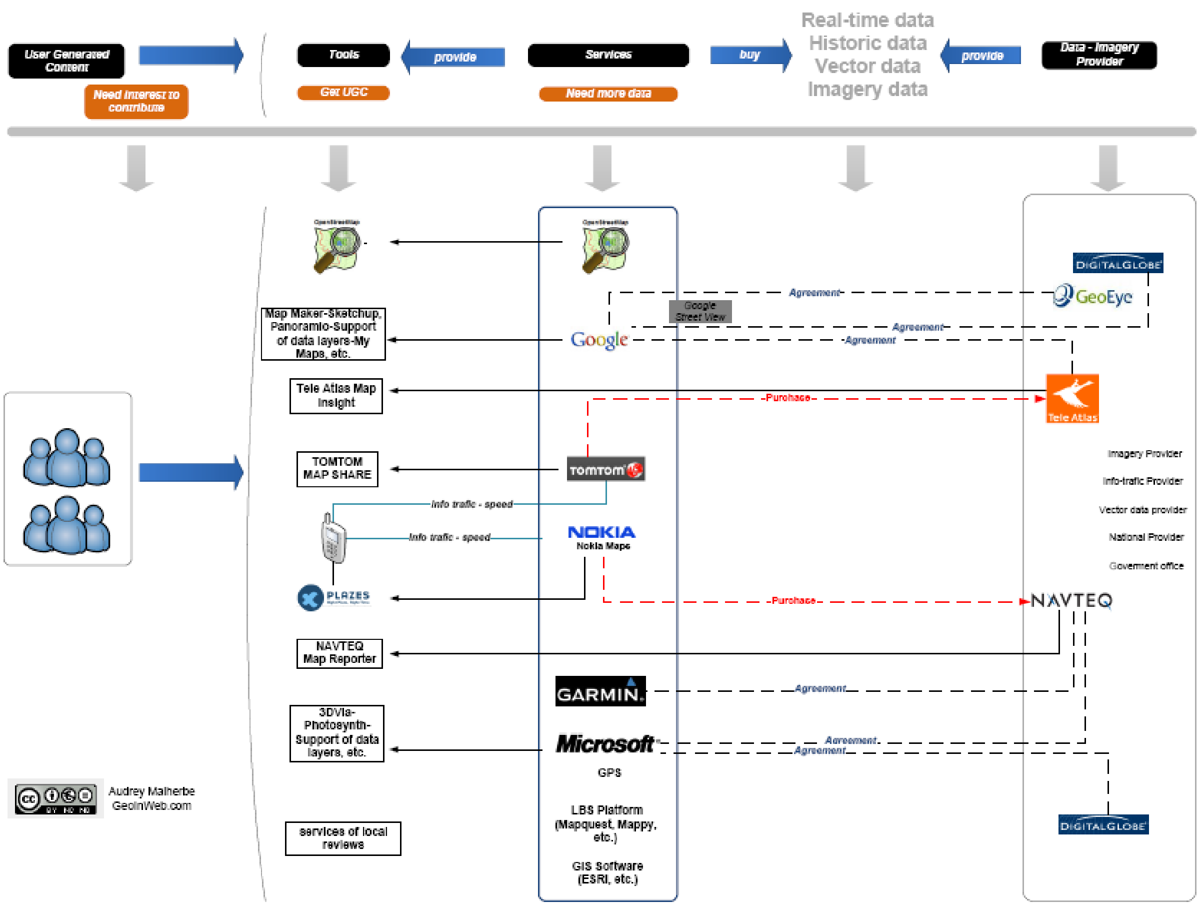


Source: http://wiki.openstreetmap.org/wiki/Stats

Figure 2.7: Graph of the users and contributions in OSM.

Commercial Map Maker

Besides OpenStreetMap, commercial mapping companies such as TeleAtlas, NavTeq and their partners saw in the VGI phenomenon opportunities to complement and replace part of their expensive and time-consuming process to generate and update data. Contrary to open data, the use of these data is restricted by the licenses of the companies and therefore contributors cannot reuse the data for their own purpose. The only incentive for contributors is in the form of better services that they can get in return from these companies thanks to the improved data. Commercial software to report feedback or create new data includes Google Map Maker, Tele Atlas Map Insight, Tom Tom map share, and Navteq map reporter. Malherbe (2009) summarizes this process as illustrated in Figure 2.8.



Source: Malherbe (2009)

Figure 2.8: User-generated content to correct commercial geospatial data

eBird

The inventory of biological species such as plant specimens or birds for biodiversity studies has always been a domain where scientists need help from volunteers due to the extent of the task. Therefore, exploring VGI avenues as new ways to facilitate the collection of geographic information on species by volunteers appears like a logical process (Klinkenberg, 2009). eBird was launched in 2002 by the Cornell Lab of Ornithology and National Audubon Society to provide information on bird abundance and distribution based on the observations made by volunteer recreational and professional bird watchers. Bird observations can be submitted directly on the eBird Website through a stepwise process that includes location of the observation (by geocoding, pinpointing on a map or direct input of coordinates) and input of attribute information such as the date and the type of observation along with a free form note. The data collected are aggregated in a database and made available through interactive maps, graphs, and bar charts that can also be downloaded. As of 2006, participants reported more than 4.3 million bird observations across North America.

Other similar projects: Another very similar Website is Geobird (http://geobirds.com/). Other scientific domains that require a large number of time-consuming measurements have started using VGI such as volunteers mapping vernal pools in New Jersey (Tulloch, 2007) (http://www.state.nj.us/dep/fgw/ensp/vernalpool.htm). The Snowtweets project at the University of Waterloo uses twitter to collect snow measurements (http://snowcore.uwaterloo.ca/ snowtweets/snowbird/) and E-Flora BC which allows users to report suspected new sightings of invasive plant species in British Columbia (http://eflora.bc.ca/).

2.3.4.2 Local Knowledge – Type 2 VGI

Description

Type 2 VGI or local knowledge-based VGI is contributed by the users to give a local perspective on a particular issue or topic, or simply to provide a description of their surroundings. To do so, in contrast to Type 1 VGI, the data are different than usual GIS knowledge. They are mainly unstructured under the form of free text that enables people to express their own perceptions and ideas. Along with the free form text, users are often asked to categorise their contribution freely or according to predefined themes in order to facilitate the data browsing. A three way interaction is enabled in Type 2 VGI applications, rather than aiming at improving the data quality like in Type 1 VGI, it seeks to facilitate community discussion in order to identify areas of conflict or agreement. In some cases, the aim of the data collection is only to survey the public and interactions between the users are snot enabled. Type 2 VGI can be collected in an open volunteered process. The aim is usually to gather people's descriptions about their surroundings (see Wikimapia and Flickr example). Such data are implicitly subjective and can be used to identify places that are representative of people perception instead of standardized geographic boundaries. Type 2 VGI can also be collected around a specific issue or topic as part of a process facilitated by local governments or community organizations. These data are usually openly subjective and involve users' opinions. This case is termed facilitated-VGI (f-VGI) by Seeger (2008) and is especially relevant to involve people in local planning processes or collect reports about exceptional events (see Ushahidi and MapChat).

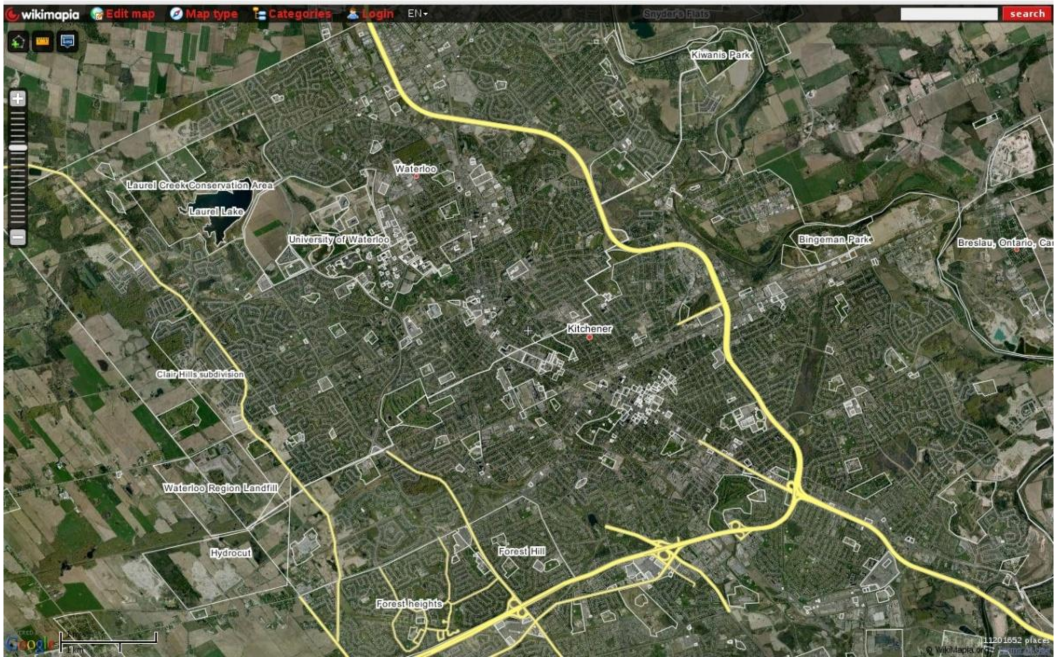
Examples

Wikimapia

Wikimapia draws upon the concepts and procedures from Wikipedia and applies them to the production of an editable and interactive map of the whole world. Wikimapia differs from OpenStreetMap in the kind of spatial features and attributes that are collected. In Wikimapia, the data contributed are focused on the identification and description of places such as counties, towns, schools, hospitals, restaurant and any building of particular interest. These places were originally identified by simple bounding boxes drawn over an aerial photography but the system has evolved to allow users to draw complex polygons and linear features (see Figure 2.9). Once a spatial feature is drawn, the feature is described by a free form description that can incorporate pictures and videos and by a category that identifies the type of feature. As mentioned on the Wikimapia wiki, the data cannot relay opinions and views and are therefore supposed to be objective. However, as mentioned earlier, even the spatial definition of the places can be subjective.

The number of places added to the map since the launch now surpasses 10,000,000. These data are licensed to Wikimapia and can be used only through their Website or API which is significantly constraining compared to OSM. However, as mentioned by Goodchild (2008), Wikimapia provides a richer alternative to a standardized gazetteer as it can show the multiple names and perspectives of the geographical places which is an important potential of VGI, as identified by Elwood (2008b, 2009) and discussed in Section 2.4.6.

Other similar projects: VGI project like CommonCensus (http://www.commoncensus.org/) (Tulloch, 2007) or the Toronto neighbourhood map elaborated by the Toronto Star (Kidd, 2009).



Note: Each white polygon represents a place that is described in Wikimapia

Figure 2.9: The region of Waterloo in Wikimapia.

Ushahidi

Ushahidi is designed to provide a simple way to gather information from citizens for use in crisis response. Such VGI can be extremely powerful as they can provide insights into crisis events that are inaccessible or neglected by the mainstream media. Ushahidi was first used in Kenya to map reports of post-election violence fallout in 2008. It has since been used to cover other events (http://www.ushahidi.com/work) such as the Indian Election, the xenophobic attacks in South Africa, and the Haiti earthquake. Practically, Ushahidi is a platform to collect data contributed by citizens through text messages, email, or a Web form input. The data input consists of a free form comment along with an optional category chosen from a predefined set. The location of a report is determined through geocoding or pinpointing on a map. Even though the data are supposed to be objective, in an emotional context like a crisis event, some data may be biased or exaggerated. Users can interact with each other by commenting on others' entries or by voting on the credibility of a report. The data collected are then aggregated and made publicly available on a Website where they can be visualized through maps, timelines and graphs. Ushahidi is only at its beta version but the collaborative open source development it uses quickly brings a lot of improvements such as a Twitter module and smart phone applications to broaden the possibilities of inputs.

Other similar projects: Similar applications have been created to cover various general events less focused on humanitarian crises. One of the first ones was Twitter Vote Report, a platform for crowd-sourced election monitoring. Twitter Vote Report allowed US voters to report voting irregularities in real-time. Another example of such an application was created by Ben Marsh (http://www.benmarsh.co.uk/) to cover the exceptional snowfall in the UK in February 2009. FixMyStreet allows residents to report local problems such as graffiti, potholes and street lightning issues. Finally, the ethically questionable RottenNeighbour allows user to report nuisance made by neighbours. However, such data are definitely subjective.

MapChat

MapChat is a Web-based tool developed to facilitate map-based discussion between citizens, planners and policy makers. The case study presented in thesis uses data that have been collected with MapChat so its presentation is discussed in more details in Chapter 4. To enable a map-based discussion, MapChat allows users to draw features on a map (points, lines, polygons) or select existing features and to leave comments linked to these features or drawings. The comments are unstructured free text. Users can also respond synchronously or asynchronously to other people's comments to agree, disagree or simply add details to map-based discussions. MapChat is used in a facilitated process where a government agency or an NGO can use the software to collect input from citizens on a specific topic (see Hall et al., 2010).

Other similar projects: Many other examples of applications for collecting input from local people can be found in the PPGIS literature (Craig et al., 2002a; Seeger, 2008). One of the earliest Web-based applications is Virtual SlaithWaite (Carver, Evans, Kingston, and Turton, 2001; Kingston, 2002). It allowed users to select features on a map a leave comment on them. However, users could not interact with each other and they could not see the resulting dataset.

Flickr

Flickr is an online photo management tool where users are able to upload, edit, organize, and share pictures. Photographers can add a free form description and descriptive tags next to each picture. The free tagging method gives some structure to the alphanumeric component of the data and facilitates browsing. Pictures can be located on a map by dragging them to a specific location. Other more advanced methods involving third party tools and GPS unit can be used. The most popular method is to use the EXIF format which allows users to embed the coordinates of where the picture was taken in each picture file. Flickr can then directly read the EXIF tag from the files and automatically position the pictures on the map.

Anyone is free to join the Flickr community and can share or keep private the pictures they uploaded. When the pictures are shared, users can interact and comment on other people's pictures. To date, over 3.6 billion pictures have been uploaded to Flickr and about 83 million are geotagged at a pace of about 3.5 million a month. This large amount of VGI made of free descriptions and pictures offers researchers new avenues to study people's perspectives on places (Dykes, Purves, Edwardes, and Wood, 2008; Purves et al., 2009; Purves and Edwardes, 2008). It is also important to note that picture sharing platforms can also be used only for personal interests to share privately pictures with relatives. To reflect these different uses, these applications are located in between LK and PK.

Other similar projects: Panoramio and Ipernity are other projects almost identical to Flickr. Panoramio is owned by Google and therefore it has the advantage of being able to exploit Google Map data and the associated API. Ipernity extends the sharing concept beyond the photos with videos and audios.

Yelp

Presentation: Yelp is a Website that provides a listing of businesses like the Yellow Page combined with a platform allowing users to leave reviews on the businesses and services. Each business is sorted in one or more categories and is located on the map thanks to the geocoding of their addresses. Users can leave free form reviews accompanied by a rating on a scale of one to five. They can also interact with each other by leave reviews and ratings as well as through other social networking features like direct messaging and online communities. Recently, in response to the criticism stating that businesses were defenceless against negative reviews and also that some of the reviews may be wrong, Yelp added a feature to allow business owners to reply to a review. The contribution process is based on volunteering and the incentive for users to contribute to share their good and bad experiences.

Other similar projects: Dismoioù is a francophone equivalent to Yelp. EveryTrail deals with another topic as it allows users to share their trips or hikes. Users can upload a GPS track of their hike that is then represented on a topographic map. Furthermore, users can upload the pictures taken during the trip to get them georeferenced along the trail at the place they were taken. Since this kind of VGI applications often offers social network features in order to share experiences with friends, they are located at the transition between Type 2 and Type 3 VGI.

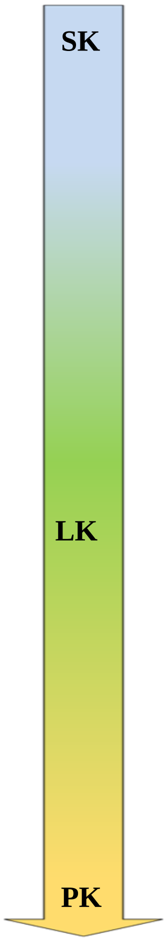
2.3.4.3 Personal data – VGI Type 3

Description

Type 3 VGI is produced using a third group of applications, inspired by social network applications, which allow users to share georeferenced data with their friends and relatives or sometimes with anyone. These applications are sometimes also termed location-based social networks. This type of VGI has been popularized by moderately cheap smart phones allowing for automatic geolocation through the use of the phone or Internet networks. The primary functionality is that users can share their position with their friends and vice versa so that they can get in touch with them when they happen to be nearby. Users can also share georeferenced stories, pictures and recommendations and in the other way around they can get information on their immediate surrounding about point of interests or events that has been previously posted by friends or even any users. All the information exchanged within these social networks is mainly subjective and unstructured. It is usually kept private, especially the user's location but some information such as events might be shared with everyone. Such applications bring questions around privacy. Users usually are in total control over what they choose to share, however an error in handling the software can easily happen or even information might be collected in a process unbeknown to the users.

Examples

BrightKite, Loopt, Plazes, Dopplr and others (see Euvrard, 2009) have a few differences but they more or less allow users to perform the actions described above. Some applications such as Foursquare and Gowalla use the geolocation and social networking aspects as a base for games. The solution developed by Google called Google Latitude is focused only on enabling the user to share his location and locate his friends. Fireeagle is slightly different from the example cited above because it is platform that allows developers to integrate users' location in applications they build (examples include Brightkite and Dopplr).



Application	Data				Process	
	Spatial		Attribute			
	Capture	Feature	Structured/ Unstructured	Objective/ Subjective	Interaction	Nature
OpenStreet Map	GPS tracks Digitizing	Point, line, polygon	Structured	Objective	n way	Volunteered
eBird	Geocoding Pinpointing GPS	Point	Mainly structured	Objective	2 way	Volunteered
Wikimapia	Digitizing	Bounding box and polygon	Unstructured (free description) Structured (category)	Mainly Objective	n way	Volunteered
Ushahidi	Geocoding Pinpointing	Point	Mainly unstructured (free text) Structured (category)	Objective/Subjective	n way	Volunteered
MapChat	Drawing Selection	Point, line, polygon	Unstructured	Objective/Subjective	n way	f-VGI
Virtual SlaithWaite	Selection	-	Unstructured	Subjective	1 way	f-VGI
Flickr	Pinpointing GPS	Point	Mainly unstructured Structured	Mainly subjective	n way	Volunteered/Kept Private
Yelp	Geocoding	Point	Unstructured and Structured	Subjective	n way	Volunteered
BrightKite	Networks/ GPS	Point	Mainly unstructured	Subjective	n way	Kept private

Table 2.4: Review of a VGI sample

2.3.5 VGI Type Addressed

As presented in the previous section, the term VGI does not refer to a single type of data but rather to multiple types of data with different characteristics and contribution purposes. As a consequence, research questions and challenges vary according to the different types of data. On the one hand, some challenges are common to all three types of VGI with different emphasis depending on the type. For instance, the issue of credibility and quality of volunteered information is relevant for any VGI type. However it is particularly crucial for SK. On the other hand, some challenges are specific to certain VGI types, for instance the issues around privacy with PK. Given that it is beyond the scope of this thesis to evaluate potential solutions to all of these challenges, the remainder of the thesis focuses on Type 2 VGI that raises interesting challenges. As mentioned in Section 2.3.3, Type 2 VGI is greatly different from most usual geographic data. It involves new forms of knowledge conveying perceptions and opinions with data that are subjective and unstructured. Only a few methods and tools have been designed to deal with this type of data, whereas Type 1 VGI can more easily draw upon existing tools designed to deal with standard geographical data.

Furthermore, beyond the simple handling of such data, the analysis and exploitation of perceptions and opinions embedded in the data can be valuable to describe places (Purves and Edwardes, 2008) or can be integrated in a planning or design process. However, in their raw form, these data have sometimes very little usability for planners and decision makers especially as they are increasingly generated in large volumes. As stated by Tulloch (2008, p. 168):

> "[..]highlights the need for planners and designers to become more familiar with these technologies and various forms of VGI to that they are properly equipped to find the proper meaning in the cacophonous deluge of VGI that is becoming available."

Planners and the broader public must become trained to use such data and new tools are required to improve data browsing and visualization that enables meaning to be extracted. Such tools also enable a better collaboration between the users who can easily browse existing data and therefore contribute data that are more relevant (Hopfer and MacEachren, 2007). Citizens or community-based organizations would also be better equipped to use these data to build cases to defend their causes. The remainder of this study focuses on Type 2 VGI and especially on improving its usability through visualization techniques.

2.4 The Power of VGI, its Related Challenges and Possible Solutions

2.4.1 Introduction

The major differences between VGI creation processes and regular geospatial data creation processes give VGI unique characteristics, as noted by Elwood (2008a, p. 133):

> "These technologies and practices are dramatically altering the contexts of geospatial data creation and sharing, the individuals and institutions who acts as data producers and users, and perhaps most strikingly, geospatial data themselves."

As such, using VGI raises new challenges as well as it stresses old cartographic challenges such as the representation of vagueness (Fisher, 1999). These challenges significantly limit the usability and the utility of VGI and they therefore need to be addressed to take advantage of the full potential of VGI. However, using VGI requires the adaptation of existing concepts and tools as well as the creation and experimentation of new techniques. A review of the main challenges related to VGI and especially VGI Type 2 is presented in the next section.

2.4.2 Quality, Accuracy, Credibility

The issue of the quality of VGI has attracted most attention to date. Similar to what happened in the journalism field where blogs were criticized by professional journalists for neglecting established journalistic practices, some professional map-makers feel threatened by the emergence of amateur mapping and question the quality of VGI (White, 2008). There are several legitimate criticisms that stem from the fact that VGI changes dramatically the geospatial data creations and dissemination processes.

Standard geospatial data creation processes are characterized by the presence of professional gatekeepers who act as an authority by controlling data quality and maintaining standards. However, the Internet and the Web 2.0 lower the cost of creating and disseminating information and consequently create an abundance of information produced collaboratively by multiple users, most of whom typically remain unidentified. In this context, the common concept of authority embodied by gatekeepers is no longer viable (Flanagin and Metzger, 2008). Furthermore, users' skills are very disparate and tools that are supposed to embed the required expertise may have built-in errors that impact data quality such as imagery that happens to be offset in Google Maps (Goodchild, 2007).

Therefore, the concepts of credibility and quality of the information based on the established expertise of the creators are no longer applicable. Finally, the information might be manipulated with some forms of “geospamming” (Sieber, 2007) or politicised by certain users. Overall, these concepts, fundamental in the regular data production process, need to be re-worked and adapted in the VGI context in order to find solutions to evaluate the quality of the information. Otherwise, VGI will not be used (Maué, 2007) or could lead to serious scientific, social, personal and political consequences (Flanagin and Metzger, 2008).

In the case of Type 1 VGI, quality and accuracy is crucial for the data to be useful and adopted by scientists and users at large, as mentioned by Klinkenberg (2009) and formulated by the eBird team on their Website (http://ebird.org/content/ebird/about/ebird-data-quality):

> “A database is only as good as its weakest record. If even a few records can be deemed questionable, then the entire data set can be labelled as such”.

Thanks to its similarity to standard geospatial data, it is easier to use mainly classical methods to ensure or assess the quality of VGI. In this case, the definition of quality includes characteristics such as lineage, attribute accuracy, positional accuracy, temporal accuracy, completeness and semantic accuracy. However, quality control and assessment are rare and often not performed by the VGI communities themselves or other ‘official’ agencies for that matter. For instance, despite its popularity and its impressive amount of collected data, OpenStreetMap has no systematic quality control and assessment and only a few external quantitative (Haklay, 2008, 2009) or qualitative assessments (Michaud, 2008) have been performed. These report a heterogeneous geographical coverage, accuracy and content.

An exception is eBird that implements systematic quality control for each record based on automated data filters which detect and flag unusual records (rarity, out of season...) that are then reviewed by scientists and expert birders. Such methods can be combined with formal user trainings when the initiative has its roots in a local and materialized project (Tulloch, 2008), online resources, or spontaneous user community meetings such as OpenStreetMap mapping parties where new users are taught how to use a GPS and edit geospatial data (Perkins and Dodge, 2008). Another option is to provide materials that embed some of the required expertise such as sensors (Gouveia and Fonseca, 2008).

When dealing with Type 2 VGI, parameters such as lineage or accuracy are difficult to ascertain given the subjective and qualitative nature of the data. Indeed, such data cannot be objectively judged good or bad but they can rather be judged useful or relevant for others. Flanagin and Metzger (2008) note that it is more relevant and useful to consider the notion of credibility based on trustworthiness rather than one based on accuracy. To apply this concept, several methods can be used. A first group of methods relies on exploiting a degree further the core concepts of the VGI phenomenon (the users and the technology). Even though they are at the origin of the challenges, their power can be harnessed to design potential solutions. First, the ability for the users to comment or correct errors in others' entries allows for an ongoing quality assessment. This "self-correction" of the data becomes more efficient as the number of users increases. In addition, more formal evaluations can be implemented by allowing users to rate the quality of each contribution and the user's overall reputation (Maué, 2007) as is common in commercial Websites such as Ebay or Amazon. This creates a peer-to-peer or crowdsourced credibility assessment. The user reputation approach can be combined with a system of user levels where users have the right or not to perform some actions according to their reputation and experience, like in Wikimapia where there is no need to be registered to add a place to the map but it is required to be a trusted user to perform other actions. These approaches can also be used with Type 1 VGI to infer the reliability of a user and the credibility and quality of his/her contributions.

Such systems are usually effective from the users' point of view, as the contributed evaluations are proven to influence the credibility judgement of other users (Flanagin and Metzger, 2008). However, systems solely based on crowdsourced methods may suffer from biases due to real life social relationships (Maué, 2007) or group thinking effects (Flanagin and Metzger, 2008) that can introduce bias. Therefore, these systems can be complemented by taking into account other objective and external parameters by the use of algorithms and automated methods.

An interesting example is provided by Swift River, a project under development alongside Ushahidi. Swift River is designed to make sense of the massive amount of information collected by dealing with two of the main challenges that VGI platforms are facing, namely the information overload and the information quality. To do so, Swift River implements a two step filter that assesses the veracity and level of importance of each piece of information to validate eventually the data and establish facts. Step one consists of applying machine based algorithms to the incoming flow of data such as natural language processing to parse the content of the submitted reports and identify reports that are related to one same event to aggregated and assumed to have a veracity degree (Meier, 2009). The second step is a decentralized human filtering process that engages self-interested citizens who curate the information (Hersman, 2009).

Once the information has gone through the two step process, it is then available to be used by other applications. Other examples include a reputation model implemented by Bishr and Mantelas (2008) based on peer ratings of information, the number of times an entity has been reported, the distance between the user and the contribution he made, and the user social network characteristics as well as an entirely algorithm-based method presented Mummidi and Krumm (2008) to identify points of interests by finding geometric clusters of contributions and examining their texts. These three examples rely on the number of contributions and relate the number of similar reports with their quality. However such a method is difficult to apply when VGI is collected in small amounts. Also, it can create a bias between popular and rare knowledge or between highly and low populated areas.

Methods based on reputation and/or algorithms are good alternatives to regular quality assessment and control in order to assess the relevance of the contributed data. Moreover, these methods can serve to deal with visualization issues due to information overload as described in Section 2.4.5. However, the notion of relevance can depend on the person looking at and using the data. Two people in two different contexts may want to look at different aspects of the data. In this context, it seems better to consider filtering techniques that can help a user find and refine the information that he/she deems relevant as noted in Section 2.4.5.

2.4.3 Heterogeneous

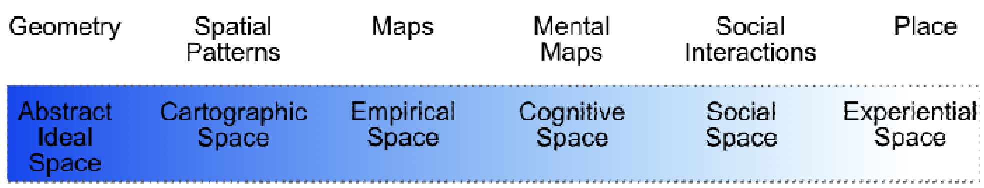
The users' perspectives, skills and vocabularies get embedded in VGI, especially of Type 2. Given the extreme variety of people who contribute data, resulting datasets are likely to be highly heterogeneous (Elwood, 2009). Moreover, the heterogeneity tends to increase as the volume of data increases. For instance, within a dataset, similar words can be used to express different ideas or different words can be used to express the same ideas. Similarly, accuracy can vary from one feature to another according to the methods employed for data input, the user skills and knowledge, and even the scale that a feature is drawn at. Data can also be heterogeneous in terms of space. Some areas may be replete with VGI data whereas others may be almost empty of any users' contribution. Such heterogeneity can be interpreted to identify areas of particular interest. However it has to be interpreted carefully. Indeed, spatial variations in the data can be due to other reasons such as high and low population densities or inequalities in access to new technology. Overall, the observation of VGI data density can be considered to provide insights as described and used in Chapter 3. Moreover, methods to visualize the spatial variation of information density can also be used to facilitate the data browsing as described in Section 2.4.5

Heterogeneity of VGI is also problematic when integrating different datasets together. For instance, different VGI applications may address similar issues but with different perspectives or different semantics. Approaches to tackle this issue can rely on the standardization of terms, the use of ontologies for automatic integration or the use of metadata for a more manual integration (Elwood, 2009). However, such standardization seems hard to apply for the Type 2 VGI that is volatile and unstructured. In this context, studying the dynamic evolution of the data may yield more interesting insights as described in Section 2.4.6. Similarly, the integration of VGI data generated in a bottom-up approach with usual geographic data created in a top-down approach is a challenging issue. However Goodchild (2007) sees an opportunity for VGI to be integrated in with regular data sources because of the decline of mapping agencies. Such integration would be based on the emergence of a patchwork model.

In this model, VGI complements the official data infrastructure by providing localized data collection that helps generating the complete data coverage. Such an approach seems entirely appropriate for Type 1 VGI with, for instance, Web sensor applications (Craglia, 2007; Gouveia and Fonseca, 2008). However it seems a lot more difficult to put into practice for Type 2 VGI that include new forms of knowledge (see in Section 2.3.3.2). Thus the best way to take advantage of Type 2 VGI may not be its simple integration with official data. Instead, the power of Type 2 VGI may lie more in the exploration and exploitation of its richness (qualitative, multiple view points, dynamic...) to bring different viewpoints and knowledge to a community and eventually be integrated in formal decision-making processes. Such an approach can be based on the use of analytic and visualization tools, as mentioned in Section 2.3.5 and discussed further in the reminder of this section.

2.4.4 Qualitative and Subjective Knowledge

As presented in the previous sections, the qualitative and subjective nature of Type 2 VGI raises issues on the evaluation of its accuracy and quality and on its integration with regular data sources. However, these aspects of Type 2 VGI are also an important source of its richness. Indeed, as raised by GIS critiques in the 1990’s (see Section 2.1), official GIS data offer a restricted view of the world and fail to integrate local and subjective perspectives that can be as, or more, valuable. In this context, VGI provides a great source to complement official data by allowing users to express the world as they perceive it through their rich and vague viewpoints. However, the use of VGI with regular GIS is complex. Interpersonal discourse is inherently vague and semantically rich, whereas GIS data models are simple and precise (Goodchild, 2002). Therefore, Type 2 VGI is difficult to integrate in regular GIS data models. As mentioned in Section 2.4.3, the simple integration and use of Type 2 VGI with existing tools and data is likely not the best way to exploit Type 2 VGI. Indeed, as conceptualized by Purves and Edwardes (2008) in Figure 2.10, standard GIS data or Type 1 VGI and Type 2 VGI do not meet the same needs. Standard GIS data are useful to represent abstract space which is suitable to perform analyses, whereas Type 2 VGI can be used to describe places based on human perceptions. To exploit this VGI richness, existing methods and tools need to be adapted and reinvented. For instance, Wikimapia (Goodchild, 2008), CommonCensus (Tulloch, 2007), the Toronto Star neighbourhood map (Kidd, 2009) or some exploitation of Flickr data (Catt, 2008; Dykes et al., 2008; Straup Cope, 2008) illustrate how the richness of VGI can be used to identify and visualise areas that are representative of people's perception of places rather than standardized geographic areas.

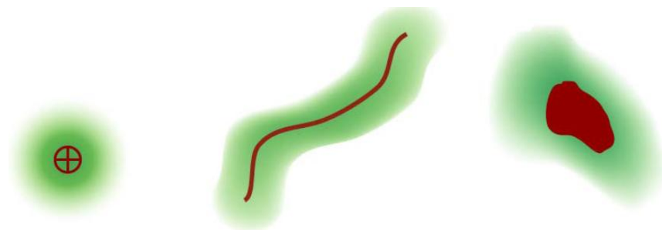


Source: Purves and Edwardes (2008)

Figure 2.10: Space-place continuum

Another aspect of the qualitative facet of VGI concerns the vagueness of spatial objects in people's representation, as mentioned by Kingston (2002, p. 111): "people's everyday life involves fuzzy entities which are not bounded by neat lines". This topic can be seen as similar to the long standing issue of representing uncertainty in GIS (Fisher, 1999; Zhang and Goodchild, 2002). However, there are few practical implementations and experimentation of the uncertainty or vagueness representation as it was not major in common GIS needs. In the context of VGI, De Longueville, Ostländer, and Keskitalo (2009) started to implement a system based on an Open Gazetteer approach and the concept of Degree of Truth. The Open Gazetteer approach is similar to the Wikimapia principles but focusing on a small area of study.

In parallel, the Degree of Truth allows for the representation of vague spatial objects by using a membership function as shown in Figure 2.11. The membership function is based on user-encoded vagueness assessed by the contributor and system-created vagueness metadata based on the zoom level at which the feature was drawn. Such a practical example is a necessary first step to represent the real nature if VGI and support the credibility of VGI.



Source: De Longueville et al. (2009)

Figure 2.11: ‘Egg-yolk’ representation

To allow users to express their complex thoughts, the bulk of Type 2 VGI is collected in a free text format linked to spatial features. This free form text can contain facts, opinions, perceptions, and references. Such semantically rich data are complex to handle, analyze and represent with traditional GIS tools. For instance, such data are difficult to classify into well defined themes that would permit standard GIS layers to be created. To address this issue, VGI applications can attempt to design categories beforehand and then ask users to select a category besides their free text. This method can provide an easy mechanism to classify data for VGI applications having a well defined content. However, for more flexible applications it can be relevant to use a “folksonomy” system as described and used in Chapter 3. Similarly, other data format that are linked to spatial features such as pictures and videos raise important challenges, as mentioned by Craglia (2007, p. 3): “Other strands that are imagery based (photos, movies, blogs or annotations) pose good research challenges in how they can be searched and documented, and in particularly how they can be harnessed to contribute to analysis and informed decision making”.

Another example of the richness that can be exploited in VGI data are the spatial references contained in the free text linked to the spatial feature. Indeed, a user comment is usually linked to one spatial feature but a user may also explain how this feature spatially relates to other features. Such references can be exploited to clarify the message for other users (Rinner et al., 2008) and to give insights on the user’s perception of spatial relationships. For instance, these references could be spatialized automatically and be displayed on the map as secondary spatial features linked to the message. Further research on the spatialization of VGI can draw upon the work done to spatialize or geoparse Webpage content (Scharl, 2007), or books such as the project Gutenkarte (http://gutenkarte.org/) which spatializes literature classics.

Finally, it seems that one of the most appropriate approaches to explore the qualitative richness of VGI is to use visualization tools in combination with other methods. For instance, aspatial visualization tools such as tag clouds or tree maps can be used to visualize the complex VGI semantic (Purves et al., 2009). Spatial methods can be used to visualize information density (Dykes et al., 2008). Furthermore, these approaches are complementary with the need to explore, browse, and sort VGI content and get meaning from it.

2.4.5 Browsing and Filtering

Opening the data creation to many users is extremely powerful to collect massive amounts of data in a short time as illustrated by the mapping of Haiti through OSM in the aftermath of the earthquake in January 2010 (ITO World, 2010; Maron, 2010). Important amounts of data are usually manageable when dealing with Type 1 VGI as they are structured according to usual GIS models. However, when dealing with Type 2 VGI, the lack of structure makes the data difficult to browse and important amounts of data often result in an overload of information for the users. In these conditions, the utility of the data is limited since they are difficult to explore, sort and filter. Information overload can occur in terms of content when there is redundant information or simply because there are too many contributions to read. In terms of visualizing contributions on a map, the spatial features may be cluttered and the user cannot distinguish one feature from another.

In these cases, filters and aggregation methods based on the evaluation of the data quality can be used to reduce the information overload. Information deemed erroneous can be discarded and duplicate data can be aggregated through automatic and human-based approach as described by some examples in Section 2.4.2. Other methods can be based on the notion of “interestingness” rather than quality. For instance, the Flickr labs that look to improve the exploration of the millions of pictures, implement a measure of “interestingness” (http://www.flickr.com/explore/interesting) based on many parameters such as the comments on the picture and their origin, the ratings, the tags, the number of views and other parameters. However, such methods simplify and aggregate the data in ways that may not be acceptable or possible in certain cases.

First, in political and social contexts, the data collection and processing phases have to be transparent for the data to be trusted and used by people. Indeed, applying algorithms may seem to be a black box system that certain users will not want to trust. Therefore in this case, the process behind the simplification of the raw VGI data must be open and easy for non-experts to interpret. Also, these algorithms tend to favour popular areas since they often base their calculations on the number of similar contributions. In certain cases users may want to have access to all the data and especially to contributions that are relatively unique as they may illustrate the opinions of communities that are isolated geographically or technologically. Moreover, such methods may not be efficient with data that are collected in amounts that have significance to local interests but are too limited in volumes to make the algorithms meaningful. This critical quantity is often not reached by Type 2 VGI initiatives where the data creation is done by a comparatively small numbers of people within a precise context as illustrated in the case study presented in Chapter 4. In these various cases, methods of browsing and filtering the data according to users' criteria is better suited to give local populations access to relevant data. To achieve this, VGI research can draw upon advances in the geovisualization field with regard to the development of visualization tools to browse and explore structured scientific data interactively (Dodge, McDerby, and Turner, 2008a; Dykes, MacEachren, and Kraak, 2005). While many of these advanced geovisualization tools are too complex or ill-suited to unstructured Type 2 VGI, VGI research can build upon some of the simple geovisualization concepts.

One of the main obstacles to using visualization tools to explore the data effectively is the lack of structure and classification of the data. To address this issue, the data can be categorized directly by the user as they input it or afterwards by power users acting as data curators who review and categorize the data. However, in both cases there are difficulties in their practical application as discussed further in Section 3.2.4. The process of categorizing and the categories are often referred to as tagging and tags in the Web 2.0 context. Once a satisfying categorization method has been applied, it can then be used as a base to facilitate data browsing and reduce the issue of information overload with visualization tools. For instance, a category filter allows the user to choose a category and filter the spatial features displayed on the map to show only the one from the chosen category. Numerous Websites implement such solutions amongst which the Yelp and WhoIsSick (http://www.whoissick.org) Websites are good examples.

Another way of filtering the data to reduce the information overload and help the users to find relevant data is to use the spatial component of the data. Indeed, some users may be only interested in data that are close to where they live or to an area of interest, hence a spatial filter which is actually nothing more than a geographic query in classic desktop GIS can be used. For instance, Yelp implements two types of spatial filter. One allows the users to draw a box on the map and display only the results included in this box. The second one allows the user simply to filter the results according to the map extent being display. A simple spatial filter based on the extent of the map being display is implemented in Section 3.3.5.3. Another interesting example related to spatial filtering is the ability to explore objects through a fuzzy “nearby” concept. Users can randomly access other contributions based on the fact that they are nearby one of interest. In this way, users are free to explore the diversity of the content and to create their own idea rather than following preconceived ideas (see Straup Cope, 2009).

As mentioned earlier, the unconstrained nature of the contribution process can cause spatial features to become cluttered and overlap in some areas of a map as they accumulate. This is a major issue that prevents users from visually browsing and finding relevant information on a map. It can be mitigated somewhat by using filtering methods, but in some cases filtering might not be sufficient or could be more efficient when combined with other approaches.

A basic feature to mitigate this issue and help visually browsing Web maps can be simply highlighting of the features when the mouse cursor is over them, as it is done by Wikimapia and implemented in this thesis (see in Section 3.3.5.1). Such highlighting can be also be used to facilitate browsing of the data across different displays, usually the map display and the attribute display in a Web GIS context. This technique is called brushing and is a frequently used in advanced geovisualization tools as detailed in Section 3.2.4.

Another simple solution that is frequently used in classic GIS is a scale dependant display. Small and detailed features are displayed only at large scales whereas large broad areas of interest are displayed at small scales. The dynamic nature of Web mapping also allows this approach to be extended to scale dependent spatial clustering. In this case, features are grouped depending on their proximity to each other and the current map display scale, and the resulting groups are represented by aggregate symbols. Hence, when zoomed out all the data tend to be clustered in a few large clusters and when zooming in these clusters are disaggregated in smaller clusters to the point where the scale is large enough to display each feature separately. This method is further detailed and implemented in Section 3.3.5.4.

2.4.6 Dynamically Evolving Data

As mentioned in Section 2.3.3.3, VGI is generated dynamically through the cooperation of a multitude of users. It is therefore constantly evolving and changing, often in contrast to official data that are published and then updated at regular time intervals. As discussed further in this section, the dynamic evolution of the VGI content can be capitalised to improve the quality of the data or generate new insights.

First, the cooperative nature of the process enables a powerful data correction process. Indeed, errors are reported and the data may be corrected constantly where an active user community has developed. In this way, errors are fixed directly and quickly by the users. There is no need to wait for long update intervals. This whole process is similar to the spirit of open source software development summarized by the sentence “given enough eyeballs, all bugs are shallow” (Raymond, 2000). In both cases, the more users a project has, the quicker and the more efficient the detection and correction of errors becomes. Many commercial data providers are taking advantage of this VGI power to allow users to reports errors (Figure 2.8). Another promising avenue is the use of the citizens as Web sensors (Craglia, 2007; Gouveia, 2004; Gouveia and Fonseca, 2008). These possible integration methods link up with the idea of the patchwork model envisioned by Goodchild (2007) and mentioned in Section 2.4.3.

Citizen correction of VGI is especially true for Type 1 VGI that is objective. However it is less relevant when dealing with Type 2 VGI that is subjective. In this case, the dynamic nature of VGI allows the possibility to collect and sometimes confront different of points of view and record their evolution over time in order to understand the social complexity and dynamism of geographic knowledge as articulated by Elwood (2009, p. 260):

> “Users’ modifications of digital data have the potential to be a richly informative source of insights about social and political negotiations of meaning”

An example of such insight is the possibility to study the relationships of various place names according to different groups of people (for instance, locals and tourists or young and old generation) and how the names and boundaries of places evolve over time amongst and within these communities. In some cases, the way the data are edited and modified can reflect conflicts. For instance, in a non spatial context, Wikipedia has developed algorithms to detect when there is a controversy on an article that is edited too frequently by the same people. Detection of such conflicts and facilitation of their resolution is an important asset of Type 2 VGI. In other cases, the confrontation of ideas can be facilitated by a system of comments and feedback as implemented in MapChat. In this context, the number of replies within a thread can be a good indicator of controversial or popular topics. Overall, depending on the application, the information of the data evolution can be used in different ways to help users find relevant information.

2.4.7 Ethics and Privacy

The respect of privacy in the GIS field is a complex and long standing issue but this section focuses on how mapping data contributed by users can lead to privacy concerns. An obvious violation of privacy is caused by the identification of people at locations and times. A well-known non-VGI case is Google Street View that created concerns in several countries such as Japan (Kubota, 2009) and Switzerland (Bradley, 2009). In terms of VGI, the risk is especially important with Type 3 VGI where some applications may enable the access to other users' locations although they may want to share it only with their friends. Other issues can arise due to ill intentioned or incautious users who create data that can harm other people's privacy such as previously private information like secret indigenous locations intended to be shared only with people in a specific community or social relationships as in the case of the questionable Web site rotten neighbour, where users can leave reviews on their neighbours. Finally, a more subtle issue is the indirect violation of privacy that can occur with the exposition of spatial relationships such as user-identified issues that can decrease home values and insurance costs.

Answers to privacy questions can be of a legal nature when there is a privacy violation or libel but institutional and technical solutions can help prevent such mishaps. First, it is important to increase the public awareness on privacy issues and its danger such as the cybercasing described in Section 2.3.3.3 (Friedland and Sommer, 2010). The recent advent of social network applications and especially Facebook has contributed to raise the awareness on privacy issues, but in the context of geographical information the awareness has to be even greater and deeper as the location component of the data can have unsuspected and harmful consequences for privacy (Honan, 2009). While raising awareness on privacy, VGI applications have to provide tools to help safeguard and respect people's privacy, such as transparent and easy ways of setting privacy parameters, restricted access, systems of user levels, and the possibility to report abuse.

For instance, in the context of Flickr, locating pictures on a map can lead to issues of privacy. Indeed, a picture by itself might not be sensitive but the same picture located on a map can threaten people's privacy. To tackle the issue, Flickr offers two different privacy parameters, one for the picture itself and a geoprivacy parameter for the picture's location. It gives more flexibility to the users to share their work while protecting their privacy. Another example of privacy protection tools is the way most location based network applications allow users to manage their level of privacy with an option to share their exact, an approximate, or a wrong location, or no location at all.

2.5 Chapter Summary

This chapter first outlined the human and technological factors that have contributed to the creation of VGI and the emergence of the associated research field. In order to categorize the wide range of VGI applications, a VGI typology was constructed based on a set of key characteristics. Three main types of VGI were identified along a continuum. For the remainder of this thesis, the focus was placed on one type termed Local Knowledge or Type 2 VGI. A review of the advantages and challenges associated to this type of VGI has shown that its characteristics make it a unique source of data but also bring awareness of several challenges that make its utilization difficult.

Among these challenges and possible solutions, the remainder of this thesis focuses on evaluating the potential of interactive visualization techniques to deal with the issues of information overload and extraction of meaning.

Chapter 3
Software Design

This chapter discusses the design of a prototype application built to evaluate geovisualization techniques with regards to the challenges of information overload and meaning extraction. Section 3.1 presents the research field of geovisualization in relation with these challenges. Building upon the discussion from the previous chapter, Section 3.2 discusses the generic needs and issues related to the design of a prototype application that builds upon geovisualization techniques to improve the browsing and visualization of Type 2 VGI. Section 3.3 describes the development of the prototype including its general architecture and the implementation of the visualization techniques.

For the design of the prototype application, a generic structure of Type 2 VGI data is considered. Based on the VGI framework presented in the previous chapter, such a structure can be represented by two components. First, a set of basic alpha-numeric data, namely the author of the contribution (which could be anonymous), the time (i.e. when the contribution was made), and a free form text, also referred to as text note. Second, a spatial feature (i.e. point, line or polygon) also referred to as map annotation or user drawing that is linked to the alpha-numeric data. The specific VGI data used in the case study is described in Chapter 4.

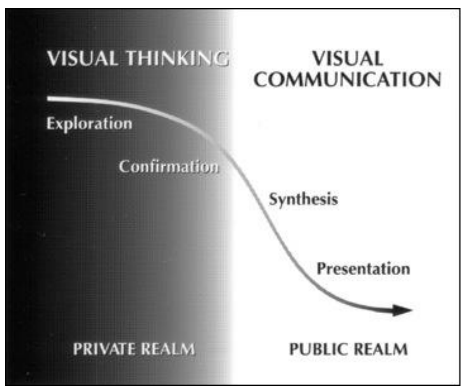
3.1 Geovisualization

Geovisualization is a powerful tool to examine and analyse datasets that are too large and too complex to be studied directly. It exploits the mind's ability to comprehend complex trends and relationships more easily in visual representations (Dodge, McDerby, and Turner, 2008b). A range of visual methods can be used at the various stages of a typical research project as presented in DiBiase's (1990) "swoopy" diagram (Figure 3.1). When studying space-related concepts and data, the visual power of maps can be used for different purposes during the various stages of the research process. These uses, characterised by MacEachren (1994) in Figure 3.2, can range from static maps, commonly used as a visual communication support for previously synthesised and chosen information, to interactive maps that can become tools to explore and analyse large and complex datasets, uncovering patterns and insights previously unknown. These two diagrams can be combined to place the different map uses within the research process as presented in Figure 3.2.

The exploration and analysis of geographic data through map visualization has developed into an established research field termed geographic visualization or geovisualization. This field draws upon a broad range of domains including cartography, exploratory data analysis and computer graphics. Geovisualization is defined by Dodge et al. (2008a, p. 2) as:

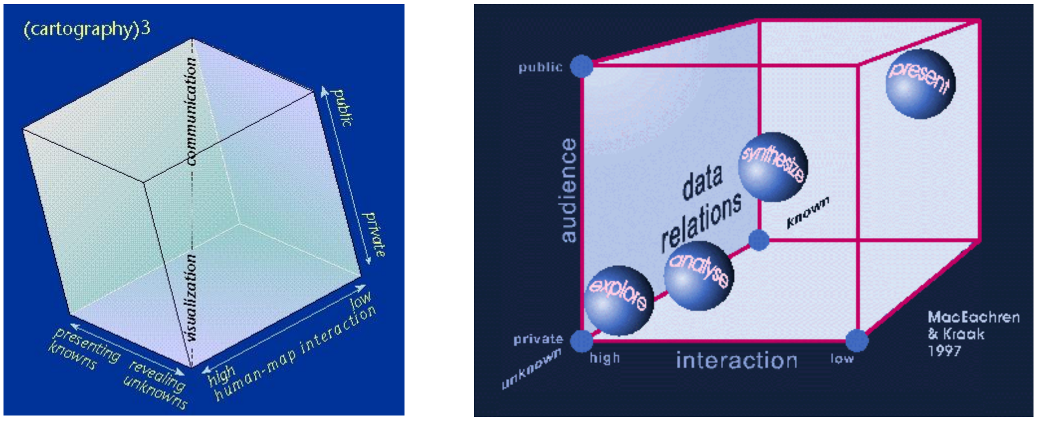
> "the application of any graphic designed to facilitate a spatial understanding of things, concepts, conditions, processes or events in the human world"

More specifically, geovisualization involves data manipulation techniques such as attribute or spatial data filtering, and data aggregation; visualization techniques such as coordinated multiple views (Keim, Panse, and Sips, 2005; Roberts, 2005, 2008; Robinson, 2009), and interaction techniques that allow users to navigate the information and to modify the cartographic representation of the data (Roth, 2009).



Source: DiBiase (1990)

Figure 3.1: The swoopy diagram



Source: MacEachren (1994) and MacEachren and the ICA Commission on Visualization (1998)

Figure 3.2: The $(\text{cartography})^3$ framework

As presented in the previous chapter, the usual lack of structure in the contribution process and the potentially infinite number of contributors can cause Type 2 VGI datasets to be very large and lead to information overload. Moreover, Type 2 VGI is heterogeneous due to its multi-authored nature and unstructured to leave enough freedom to the users which create complex data. Thus, geovisualization techniques that are useful to make sense of large and complex volumes of data appear to be particularly relevant to deal with the visualization and exploration of Type 2 VGI. Furthermore, in the context of Type 2 VGI visualization, users will have no knowledge of the data contributed by others and what it could reveal but they can discover insights by exploring the information through interactive visualization tools. However, geovisualization tools have been traditionally used with scientific data whose nature and use differ from Type 2 VGI. Thus, some methods and practices need to be adapted. For instance, they need to be accessible for a wide range of users, not only experts, and they have to be used via the Internet instead of powerful workstations and they need to deal with unstructured data. The following section discusses the generic needs related to the visualization of Type 2 VGI and the geovisualization methods that can be used to meet them.

3.2 Generic Needs and Design Issues for the Visual Exploration of Type 2 VGI

3.2.1 User-friendly

As discussed in Chapter 2, VGI is by nature created by multiple individuals. These individuals can have a wide range of skills since the new tools and techniques embed cartography and GIS expertise that makes map-making accessible to people without formal training (Goodchild, 2008). Moreover, the users operate mainly over the Internet and are therefore usually dispersed geographically and sometimes anonymously. As a result, except in a few cases (Perkins and Dodge, 2008; Tulloch, 2008), face-to-face training is not possible and web-based training presents important limitations which restricts significantly the opportunities for the users to learn complex new tools (Andrienko et al., 2002). As mentioned in Section 2.1, in PPGIS research, VGI creation can be motivated by the empowerment of individuals that are marginalized from the regular mean of communications and power, most likely possessing limited skills and access to technology.

For these various reasons, the visualization techniques employed to explore VGI must be easy to understand, use, and interpret for people with limited computer and GIS experience. However, it is possible to complement easy-to-use tools with more complex features (Nielsen, 1993) but it has to be done carefully not to harm the usability for novice users.

3.2.2 Web-based

As described in Chapter 2, the Web platform brings many advantages for PPGIS and VGI. First, it can broaden public participation as members of the public have fewer physical and time constraints to use the software. They can participate when and where they want, provided that they have an Internet connection (Carver, 2003; Sieber, 2006). The participation can be anonymous for people who are not comfortable participating in public meetings. Moreover, recent Web-based tools are generally user-friendly and they can be directly operated by users with relatively little skill.

A more technical advantage is the centralisation of the maintenance and the expertise. It lessens costs by limiting the need for experts, hardware, and maintenance, since software and data upgrades are done in one location. As a result, it also helps to address unequal access to technology, a central goal of some VGI initiatives.

The Web platform has become a standard way of contributing VGI and is on its way to become the standard way of using any software. It means that the contributors are usually dispersed and that the overall dataset is constantly evolving. Thus, to provide visualization tools for VGI that are accessible from everywhere, and present the same content for all users at any time, it is necessary to use Web-based technology.

However, two important cautions should be noted. First, even though Web mapping applications are increasingly democratized, they can still be difficult to impossible for non-computer savvy users to get started with. Even when introductory training sessions have been held, most people prefer to participate in small group contexts because it is harder to take the time and build the motivation to participate from home on your own. Second, the Internet changes the potential audience. People using the Internet are usually not the same as the people attending open houses (Von Haaren and Waren-Kretzmar, 2006). Therefore, in order to broaden the participation, it is still important to keep traditional means of communication for the data input as well as for the visualization, such as the methods presented in (Al-Kodmany, 2001). However, these approaches are not appropriate to deal with the visualization challenges related to information overload and browsing so this thesis focuses only on the Web and computer-based approach.

On the technical side, using the Web to implement visualization features has important constraints. Despite recent advances, the limited bandwidth and Web browser capabilities restrict the possibilities to develop interactive visualization tools without using browser plugins or installing rich client on the user's computer which significantly reduce their accessibility as discussed further in the software architecture section.

3.2.3 Aggregating and Filtering Data

As discussed in Section 2.4.5, the collection of Type 2 VGI leads to an overload of information for several reasons. First, the overload of information can be due to the sheer volume of data. The idea behind VGI is to build upon the power of masses. In order to maximize the amount of data collected, the contribution process is usually open to many users with no limit usually imposed on the contribution process. This leads to the creation of important amounts of data that are too large to be browsed manually. Similarly, the contribution process usually has only a very limited structure to increase the richness of the resulting datasets. The information overload is thus worsened by the heterogeneity of the contributions. In terms of spatial data, it means that contributed drawings tend to overlap and clutter in popular areas which significantly hinders the map reading and browsing. The free form text may contain various semantic fields or levels of languages which can make the data browsing very tedious; the format of the data can also hamper the data browsing. For instance, multiple geometry types may be used simultaneously and end up intertwined, reducing the map legibility. Lengthy free form texts need to be read one at a time to make sense of the data.

These issues can be mitigated by imposing more structure on the contribution process in specific VGI project (e.g. ask the contributors to categorize the data, as discussed further below) However, it is not always adequate, as it limits the potential of the contributions and often, these issues still end up occurring at different levels nonetheless. Therefore, methods to reduce the overload of information and facilitate the data browsing have to be employed. The geovisualization field and traditional cartographic principles offer several approaches to reduce the size of the dataset visualized and deal with these challenges. Aggregation and filtering are the two general sets of approach that are discussed below

3.2.3.1 Aggregation

A first set of methods consists in grouping, aggregating and assigning structure to the data in order to provide summaries of the data (Roberts, 2008). Complex knowledge discovery and data mining techniques have been used to reveal structures and patterns in datasets. The discovered patterns can then serve as a base for visualization. Sections 2.4.2 and 2.4.5 discussed some of the advantages and limitations of the first research works that have started to investigate the use of automatic processing techniques with VGI data. However, creating complex data mining algorithms was beyond the scope of the thesis and was therefore not investigated further. Instead, geovisualization methods involving limited computer processing were investigated. Even though they must simplify the data, the employed methods must allow the users to have access to the original data through interactive drill-down processes to avoid black box effects and keep a transparent process which can be essential when VGI is used in political and social contexts.

Concerning the spatial component of the data, several traditional cartographic methods that have been used to reduce the information overload in map-making can be investigated (Slocum, McMaster, Kessler, and Howard, 2008a). Except in cases of evident erroneous or abusive entries, the suppression of data should not be used as it goes again the transparency principle mentioned above. Filtering techniques presented below are an appropriate alternative to filter out data from the display. Dynamic displacement technique can be used to spread out the overlapping features as implemented in Google Earth for overlapping pictures. However, if this approach is relatively easy to implement with point data, it is significantly more complex to use with multi-types spatial data and has no practical implementation yet. Finally, inspired by classic aggregation techniques, numerous websites have started to apply clustering methods of spatial contributions in proportional circle symbols in order to speed up the display and improve the map legibility as demonstrated in the VisGets application (Dörk, Carpendale, Collins, and Williamson, 2008). The clustering is dynamically adjusted depending on the scale at which the information is available. This is particularly interesting as it enables the users to visualize different levels of aggregation or even to zoom in on to the point where he/she sees the original drawings. Section 3.3.5.4 presents in further details this aggregation technique and its implementation in the MapChat Viz prototype.

Concerning the free form text linked to the spatial component of the data, the key issue is the lack of structure or metadata that prevents the grouping of data, making it difficult to browse. Traditionally, structured descriptions of the data through attributes or metadata are created by professionals. In the context of VGI, this is no longer possible due to the huge and ever-increasing amount of data generated by a multitude of authors. Type 2 VGI data can be categorized by using various approaches. Various automatic methods can be used. The texts can be scanned directly by an algorithm to determine the number of occurrences of each word in the overall datasets and then generate a summary of the word used. More advanced natural language processing algorithms can be used to get more detailed summaries of the data.

The second approach is to take advantage of the power of the contributors to categorize the data and involve directly the authors and users in the creation of metadata, particularly categories. For instance, the Websites Delicious (http://www.delicious.com/) and Flickr (http://www.flickr.com/) have pioneered the utilisation of user-generated metadata by allowing users to categorise or “tag” Website bookmarks for the former and pictures for the latter. This method of creating metadata is coined under the name of folksonomy reflecting the “folks” origin in this taxonomy of a new kind (Mathes, 2004). A comprehensive review of the different folksonomy techniques and their values is presented in Smith (2008). Examples of approaches include leaving total freedom to the users to tag the content, dynamically suggesting tags based on the tags already used in the collection or imposing a fixed tagging structure that the users have to comply with (e.g. a set of predefined categories or a hierarchical tagging structure). As discussed earlier, it is a balance between freedom and richness and structure. The adequate solution varies depending on the context of the VGI project. Section 4.3.2 presents the tagging technique employed for the case study. As discussed in the next section, the data structure resulting from the categorization can be used for visualization purposes.

3.2.3.2 Filtering

A second way to reduce the information load is to filter out information that is not relevant to the users or in other words to enable the users to locate and retrieve task relevant subsets of information (Roberts, 2008). This is usually achieved by using interactive filtering and dynamic queries techniques that enable the users to reduce the data displayed to a subset of data. The subset of data is defined according to the various dimensions of the dataset. In the case of generic Type 2 VGI data, the standard dimensions are location, author, time, free text and the structure generated through data aggregation such as categories or clusters. The change of the filter parameters must be interactive and closely linked with the display to allow users to see immediately the results and get insights on the data as discussed in the next section. The filtering can also be enabled in conjunctive form (e.g. boolean AND) to allow the users to filter the content according to several dimensions simultaneously (Dörk et al., 2008).

Spatial filtering can be achieved by considering only the area displayed or by enabling users to select a custom area on the map with a bounding box or freehand lasso (Roberts, 2008). Free form text can be filtered by using classic word search that retrieve the data containing the searched words. The data can also be filtered according to the other dimensions of the datasets such as the time or the author. Many websites provide examples of interactive timelines used to constrain the time range of the data displayed (Dörk et al., 2008; Roth and Ross, 2009). Finally, the aggregated views of the data can also be used as a filter. For instance, users could select one or several categories to retrieve topic relevant data. The prototype developed for the case study implements interactive text and spatial filter tools described in Section 3.3.5.3.

3.2.4 Extracting Meaning

As discussed in Chapter 2, the complex nature of Type 2 VGI (i.e. multifaceted, heterogeneous, qualitative and subjective) makes its utilization challenging. However, Chapter 2 also acknowledged that data with these characteristics constitute a richness that needs to be exploited and transformed into useful information rather than being simply stored and browsed occasionally. Furthermore, this richness is even greater when looking at the dataset resulting from all of the contributions as it can reveal relationships, patterns and insights that contributions browsed individually cannot. As mentioned in Section 3.1, visual data exploration and geovisualization techniques are potential avenues to meet these needs and are investigated in this thesis. To extract meaning from the data, the users must be able to visualize interactively the multiple facets of the data (e.g. raw contributions, time, topics, spatial repartition...) and their relationships in various forms.

To achieve this goal, coordinated multiple views (CMV) techniques developed in the geovisualization field are investigated (Keim et al., 2005; Roberts, 2005, 2008; Robinson, 2009).

CMV techniques are used to build exploratory visualization environments that enable the users to look at the data through multiple interactive views that are coordinated together. A view represents one way of presenting the data. A view can have various forms such as charts, tables, maps or diagrams that can enable a variety of analytical capabilities (Roberts, 2008). Applied to the case of generic Type 2 VGI described earlier, two basic views are necessary to display the raw data, namely a tabular view to display the free from text and other alpha numeric data (i.e. author, time) and a map view to display the spatial drawing contributed. Beyond these two necessary views, several other views can be implemented to deal with the specific challenges related to Type 2 VGI.

Building upon the aggregation methods presented in the previous section, two views can be designed. First, to help users to obtain an overview of and insights on the content of the multiple unstructured text contribution, the categories can be visualized with a chart view. One popular tool to visualize categories on the Web is the tag cloud. An alternative is the simple category list ordered by the number of category occurrences. These two types of views are implemented in the MapChat Viz prototype and are further discussed in Section 3.3.5.5. Second, beyond facilitating the data browsing, the dynamic spatial aggregation of the drawings can be used to provide a view of the heterogeneous spatial repartition of the data at various scales and for various areas. This view is implemented in the MapChat Viz prototype and is further described in Section 3.3.5.4.

The other important components of CMV techniques are the user interactions with each view and their coordination. In this way, the user can look at data in one view and quickly find and analyze that same data from another angle in another view. As a result, the user can identify relationships, outliers, clusters and other patterns that are difficult to detect otherwise (Robinson, 2009). Such interaction is usually implemented through the combination of two techniques called linking and brushing, as demonstrated by many geovisualization software such as CommonGIS, GeoDa, and GeoViz (Keim et al., 2005; Roberts, 2008). Brushing is a quick and interactive selection process that lets the user select an object in one view usually through either mouse over or mouse click; the selected object is instantaneously highlighted across all the other views through linking. The highlighting is typically achieved by a changing the colour of the selected object across all the views. Other highlighting methods based on depth of field modification or leader lines can be implemented but they have yet to be widely used (Robinson, 2006, 2009). The MapChat Viz prototype implements some brushing and linking techniques between the several views (i.e. map, text, tag cloud, tag list, and spatial cluster) that are further described in Section 3.3.5.2. Beyond the simple selection through brushing, the interactive filtering techniques described in Section 3.2.3 can also be used to select subsets of data and highlight them across the different views. It allows the user to play with the filter parameters and see the changes across the several views. For instance, the MapChat Viz prototype highlights on the map all the drawings related to a selected topic.

The next section describes the software architecture and software solutions that were chosen to implement some of these techniques into a prototype designed to perform usability testing with a cross-sections of VGI users.

3.3 Software Architecture and Design

3.3.1 Goal and General Requirements

In order to provide an environment to evaluate some of these visualization techniques, a prototype called MapChat Viz was developed. As such, MapChat Viz was not meant to support the contribution of VGI or to be a decision support application but it was designed as a test bed to evaluate the relative merits of several data browsing and visualization techniques which could ultimately be implemented in other collaborative mapping software that seeks to accumulate VGI.

Several of the geovisualization techniques presented above, as well as other innovative approaches, have been implemented in existing Web applications to aid spatial data exploration. However there has been minimal testing to provide evidence that these techniques are useful and usable for the users and that their development is not simply guided by technical possibilities or individual developer choices. Moreover, the implementations of the various techniques are spread across several software applications that have each been designed for specific purposes and users. These specific contexts of use significantly influence the users' experiences and subsequently the evaluation results, which makes evaluating and comparing tools challenging. To tackle these issues, MapChat Viz was designed as a unified Web mapping tool that brings together several of these existing VGI browsing and visualization methods and enable users to test and compare them.

To achieve this goal, MapChat Viz had to meet some general requirements. First, it had to be built with a flexible development platform that permitted rapid prototyping by integrating out of the box features with customised tools. MapChat Viz required Web mapping capabilities to display generic VGI such as user's map annotations and linked free text notes over a background base map or imagery. To support the evaluation of the tools, MapChat Viz needed the ability to collect data on the usage of the features through software logging as well as an interface that facilitate users' testing and limits confounding effects by revealing different geovisualization tools progressively. Such an interface also facilitated the use of the prototype in conjunction with a workshop script and a questionnaire to collect users' feedback.

Despite the fact that it was developed to be a prototype and test bed, MapChat Viz still focuses on the participatory nature of VGI projects and, as such, targets various groups of users including a main core constituted of small community-based organisations for which Type 2 VGI is particularly interesting to publicize aspects that are neglected by official datasets as discussed in Chapter 1 and 2. This means that MapChat Viz had to be accessible for organizations which have limited financial resources and for users ranging from neophytes to experts and thus, facilitate the possibilities to reuse and build upon the tools that were evaluated. Therefore, even though not a generic requirement to visualize VGI, the decision was made to use free and open source software to enhance the dissemination of the software and the public participation as described in further details in the following section.

3.3.2 Open Source Software

The pros and cons of open source software (OSS) versus proprietary software have been extensively reviewed in the computer science literature. Therefore, this section only gives a brief overview of the reasons making OSS a relevant choice when working with community-based organisations (CBO) and PPGIS to collect Type 2 VGI.

The element that lies at the core of the OSS definition is the availability of the source code to the public, which enables anyone to run, study, modify and distribute the source code freely. This definition has many variations especially concerning the possible interaction between open source and proprietary software. These different definitions are legally formalised by different licenses. A comprehensive list can be found on the Open Source initiative Website (http://www.opensource.org/). A major consequence of the open source model is the availability of the software at no cost. It is therefore an appealing alternative to proprietary software for CBOs who often lack financial resources and cannot afford the cost of proprietary GIS software. However, the price tag argument has to be taken carefully as licensing costs are not the only costs of acquiring and deploying software. Indeed, the total cost also includes the cost of training, maintenance, extra development, and support and is usually referred to as total cost ownership (TCO). The comparison of TCO between open source and proprietary software has no simple answer and depends on the context (Holck, Pedersen, and Larsen, 2005). However, in the context of CBO, the low acquisition cost is essential to start experimenting and slowly building capacity with volunteers or through partnerships with universities.

Open source does not just mean free access to the source code. Cost is only one of many differences between OSS and proprietary software and it must not hide the other assets and drawbacks of OSS. OSS has a different approach of the development process than proprietary software. OSS applies a collaborative and transparent development process based on an open community of developers. All the communications, discussions, and decisions made by the community are public and the community is constituted based on merit. This allows for reactive software update and support as well as a good involvement of the users in the software design process. Furthermore, the contributions to the software are initially peer reviewed and then checked again by many developers as they use the code so it often provides software of better quality and higher reliability. Another important point for CBO with limited technical resources is the level of support and documentation that is available. Open Source has long been considered to belong to the hacker realm by offering poor support, documentation and usability (Nichols and Twidale, 2003). However, as an open source project becomes popular and reaches a certain level of maturity, its community gets bigger and often includes proprietary software linkages and investments. The documentation and usability then become on par with the proprietary software, and commercial support and training become available through a service oriented business model. Furthermore, the spirit of transparency, collaboration and participation of open source communities align well with the ideas of VGI and transparent planning processes carried by some CBOs. It therefore creates a synergy that can facilitate and encourage the CBO spirit of motivation and cooperation (Moreno-Sanchez, Anderson, Cruz, and Hayden, 2007).

Thanks to their low cost, their transparency, and accessibility, OSS constitutes a viable and sustainable solution for CBOs. Indeed, with the help of the researchers, they can build local capacity to take over eventually the control of the software. This would be difficult, if not impossible with proprietary software due to the prohibitive license cost and “vendor churn” (Wheatley, 2004). However, there are two important cautions for CBOs. CBOs must use mature OSS to avoid trouble such as the lack of documentation and poor usability. Moreover, IT staff or volunteers have to be highly motivated and be able to make their own way through the documentation to start. Thus, the team must contain at least a local champion or a small core group of motivated people.

The following sections described how various open source tools were used together to implement and evaluate some of the visualization techniques discussed in Section 3.2.

3.3.3 General Architecture

For the reasons discussed in Section 3.2.2 and 3.3.2, MapChat Viz was developed as a Web-based application using open source components. A typical Web application consists of three main parts, namely a client tier, a server tier and a data tier. This type of architecture is required because of the technical aspects of the development on the Internet (communication between a server and a client over the Internet network) and also because the development of applications in several tiers is a longstanding software design practice to ease the development and the maintenance of the software. This section provides an overview of the role of each tier and the communication between them with the overall architecture represented in Figure 3.3. The details of the technological choices for each tier are further detailed in the next section.

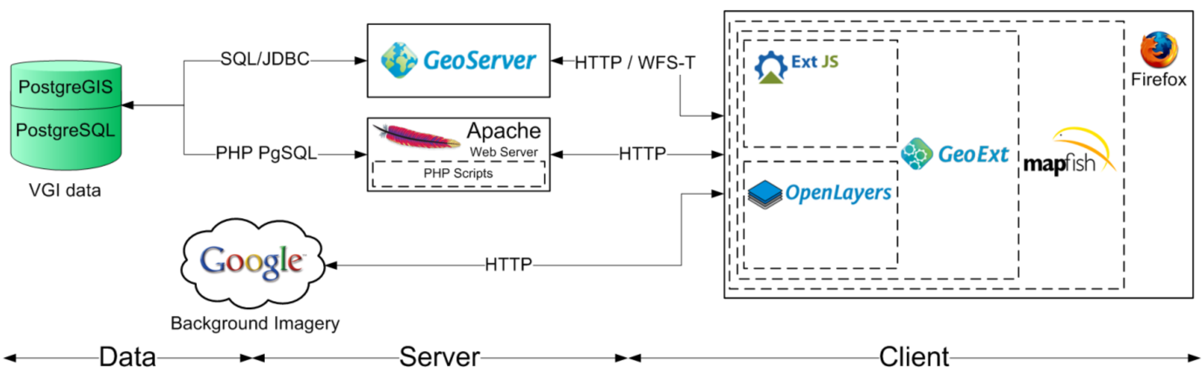


Figure 3.3: MapChat Viz architecture

The data tier handles the storage and management of the data used by the application. Two main datasets need to be handled to visualize VGI. First, the VGI data itself that consists of a vector spatial feature (i.e. the user drawing(s)) and a set of attributes including at least a text, a user id or name and the time at which the contribution was made, and optionally a set of categories. Depending on the specificity of the VGI project, other information related to the main data may be stored. The data tier must also be able to handle evolving and concurrently accessed data if the contribution process stays open while the data is being visualized. In order to manage dynamic and relational data, the solution widely used in many industries is to use a relational database system (RDBMS). In the case of VGI, the database requires a spatial extension in order to be able to easily manage the spatial component of the data. Besides the user drawings, the application needs to supply background data to provide the users with the geographic context and help them situate their drawings. This can be done either in many ways with vector data or imagery. The MapChat Viz prototype uses imagery data that are stored and served by Google and are simply requested by the client to be displayed with the VGI. This method was chosen for its convenience of use as it provides a global coverage of imagery without having to manage and publish the imagery data, but in specific cases it may be preferable to use other data sources.

The server tier comprises a Web server and a mapping server. The Web server is in charge of publishing the HTML pages and executing Hypertext Preprocessor (PHP) scripts. Based on AJAX programming practices (Mahemoff, 2006), the application consists of one Web page that is published and loaded only once when the user first connect to the URL of the Web site. Once the page is loaded, most of the user interactions are directly handled by the JavaScript code loaded on the client side. On some occasions, when the client tier needs to write or request some data from the server, it issues an AJAX request that is handled by a PHP script on the server side. The PHP script processes the request and issues the adequate Structured Query Language (SQL) queries to write or retrieve data from the database. It then returns the data or the result of the operation to the client side. For instance, some PHP scripts handle tasks such as the verification of the login/password, the processing of requests made to populate some of the views with the content requested by the user (e.g. the tag cloud), or writing the logs of user actions in the database.

The mapping server handles the requests related to the spatial component of the VGI data which consists essentially of publishing the spatial features so that they are displayed by the client tier. There are two main options to publish the VGI data. The data can be publish in an image format (i.e. jpeg or png) that is then simply displayed by the browser or in a vector format that stores the actual geometries as a list of vertex coordinates as well as the related attributes. The image format is easier to handle by the browser as it is displayed as a regular image. However, for any interaction with the map, the browser needs to request new data from server which can hamper the fluidity and interactivity of the user interaction. Vector format is heavier to handle for the browser but it enables more interactivity as each feature can be accessed directly without issuing a new request to the server. Given the high level of interactivity required by the visualization techniques discussed in Section 3.2, the users' drawings are displayed in vector format.

The vector data are transferred over the network by using the Web Feature Service Transactional (WFS-T) interface. The WFS specification is a standard interface to request and return geographic features over the Web. WFS-T is an enhanced version of WFS adding the transactional support that allows users to write and edit geographic data. Writing and editing spatial features are necessary in MapChat Viz to get feedback from the workshop participants as described in Section 3.3.1. These standards are specified by the Open Geospatial Consortium (OGC) which is a consortium of companies, governmental agencies and universities which work together to enable spatial information and systems interoperability through the creation of common standards. The WFS interface can be used in combination with Geography Markup Language (GML) which is an XML grammar defined by the OGC to express geographical features. However, a growing popular alternative to GML is the GeoJSON, which is based on the JavaScript Object Notation (JSON) format. GeoJSON is much lighter than GML and is therefore faster for the transfer of data and easier to develop with as it uses JavaScript syntax used by the client tier. Thus, MapChat Viz uses WFS-T interface with GeoJSON to transfer spatial data between the server and client tiers.

The client tier handles the User Interface (UI) of the application as well as most of the core functions. The inclusion of core functions on the client side rather than the server side, which has typically more computing capabilities, is a general trend in the development of recent Web applications. This avoids sending data back and forth between the client and the server in order to improve the interactivity of applications. However, this practice has to be used carefully not to overload the relatively limited processing capacity of the browser. As mentioned earlier, the communication between the client and the server is done through asynchronous requests according to the technique of AJAX programming (Mahemoff, 2006). This improves the interactivity of the application by updating part of the pages without reloading the entire page visible on the user browser. The client side is built with several JavaScript code libraries that are described in the next section along with the software solutions chosen for the other two tiers of the application.

3.3.4 Technical Choices

3.3.4.1 Evaluation Criteria

As shown by the comprehensive lists on the Websites opensourcegis.org and freegis.org, there are many open source GIS (OSGIS) projects under development. It can therefore be difficult to find the ones that meet particular needs. Ramsey (2007) provides a good, albeit somewhat dated, overview of the OSGIS community, its structure and its most prominent solutions that gives a good idea of the possibilities offered by OSGIS. However, in order to evaluate more precisely the different OSGIS projects and choose suitable solutions to develop the MapChat Viz prototype, a set three types of criteria need to be elaborated. The first type considers whether the solution is open source or not, for instance, feature richness and performance (Wang and Wang, 2001). The second type of criterion is related to the maturity and vitality of an open source project. These aspects have direct consequences for the level of support and documentation quality. Finally, criteria related to the specificities of the context in which the application is being developed must be considered, for instance the interoperability with existing data or the existing programming expertise.

In the context of the development of MapChat Viz the following general constraints can be identified. MapChat Viz was developed as an experimental test bed within a short time frame and with limited resources, therefore the development had to be rapid and relatively easy. As such, it had to build upon solutions that had significant features ready to use. In terms of open source related criteria, software with an active user and developer community were chosen in order to benefit from the community-provided support and documentation. The importance of these criteria enriched by considerations on the context specific criteria is developed in the following sections for each tier.

3.3.4.2 Database Tier

The database tier of MapChat Viz had one main requirement which was to possess spatial capabilities in order to store spatial features and to perform spatial queries and processing. Given this constraint, the choice was mainly limited to two database software projects, namely PostgreSQL with PostGIS and MySQL with MySQL Spatial, as these are the two main open source databases that have a spatial extension (Chen and Xie, 2008).

MySQL and PostgreSQL are two renowned and widely used database software packages (Wheeler, 2007). However, their spatial extensions have different levels of quality. PostGIS is more mature and feature-rich than MySQL Spatial (Chen and Xie, 2008; Ramsey, 2009). Moreover, it is supported as a standard backend by most open source GIS projects, which significantly simplifies the development of the whole application. Besides, the case study data were stored in PostgreSQL/PostGIS, so it was easier to use the same software to maintain consistency and avoid potential problems of data compatibility and interoperability. Thus, using PostgreSQL/PostGIS for the database tier was the most appropriate choice. Technical details on PostGIS are provided by Chen and Xie (2008).

3.3.4.3 Server Tier

The server tier consists of two pieces of software, specifically a Web server and a map server. The overall requirements on the server side of the application are rather simple as most of the application is located on the client-side to give a fast and interactive experience to the users. The industry standard open source Web server Apache (Wheeler, 2007) was chosen to generate the HTML pages and execute the PHP scripts. Concerning the map server, it had to be able to read and write PostGIS tables according to WFS-T queries, and to publish the data in a GeoJSON format that is used by the client side software. The possibility to set up the server easily through a Graphic User Interface (GUI) would be an asset to shorten the learning curve as advanced manual configuration and optimisation are not needed. In this context, GeoServer was chosen over other alternatives such as MapServer, MapGuide Open Source or FeatureServer because of its ease of use and configuration through a Web-based GUI and its direct and easy support of WFS-T and GeoJSON (Doyon and McKenna, 2009; Quadro, 2007).

3.3.4.4 Client Tier

The client tier handles the UI and most of the core functions of MapChat Viz. Since the tool focuses on the abilities for the users to browse and visualize the map data, the client tier is one of the most important aspects of the application.

In order to avoid developing from scratch and reinventing the wheel, a Web mapping application typically uses code libraries also called an Application Programming Interface (API). An API provides the developer with a library of functions to implement basic operations. As a result, developers can focus their effort on the development of the customized parts of the application. For Web mapping applications, two code libraries are typically required as foundations. First, a regular JavaScript API is needed to provide utilities such as programmable components for the UI and functions to handle Ajax requests. Second, a mapping API is required to provide features such as map navigation tools or the display of geographic layers from various sources and formats. Code libraries can provide a more or less extensive set of features ranging from basic functions to advanced out of the box features. The libraries providing only basic functions permit development of customized and optimized applications, but more time and resources are required to develop an entire application. More complete libraries can be heavier to run and come with some visual inheritance but they enable a quicker development. In the context of MapChat Viz, which is built as a prototype with relatively limited development time and resources, it was preferable to opt for solutions that provided pre-built functions and have an active community that provides support through documentation and online communication (e.g. code examples, mailing lists).

OpenLayers was selected as the Web mapping API because it has become a leading open source project in this domain with many capabilities, documentation including hands-on examples, and a strong user community backed up by companies (Schaub, 2009). Whereas other projects such as Ka-map, Chameleon, MapBuilder have progressively slowed down their development or in some cases stopped altogether (Ottens, 2008). The popular and widespread Google Map API may be seen at first hand to be an appropriate choice for this task, however even though Google Map API is free it is not open source. Hence, it is not as flexible and extensible as OpenLayers and it has constraining terms of use. It is therefore not a useful alternative. OpenLayers has many functions to handle the required mapping aspects of the application. However, it only comes with basic UI components to let the developers choose how they want to develop it.

Many JavaScript frameworks such as jQuery, Prototype and ExtJS can be used to develop the UI. From an assessment of the possibilities, ExtJS was found to have several advantages in the context of MapChat Viz. Specifically, ExtJS offers numerous out of the box controls and components such as grids, tabs, windows. Moreover, ExtJS has been integrated with OpenLayers by the project GeoExt to provide mapping UI components such as map panels and toolbars. The MapFish Client project encompasses the GeoExt components (see Figure 3.4) into a framework that allows developers to create basic Web mapping applications quickly. Thus, the MapFish Client framework served as a base for the development of MapChat Viz on top of which customized ExtJS and OpenLayers components were developed to implement the custom features presented in the following section.

To give a complete picture, it can be noted that Flash, often used to develop web-based visualizations, was not deemed a suitable solution for this work. Although it can have some advantages to develop interactive UI and animations, Flash has several documented disadvantages for developing websites (e.g. the requirement for a plugin to be installed in the browser) and due to the proprietary nature of Flash, open source development framework and mapping API for Flash are rare and rudimentary which goes against the requirements mentioned earlier.

Source: Moullet (2009)

Figure 3.4: MapFish Client structure

3.3.5 Implementation of the Prototype Software

3.3.5.1 Browsing of Geographical and Attribute Data

As discussed in Section 3.2.4, two fundamental views are necessary to visualize the alpha-numeric and spatial components of the raw VGI data. The spatial component of VGI (i.e. user's map annotations) plays a central role in the visualization of the user-generated data as it provides the context to interpret the information contributed by the user. Moreover, usability studies have shown that Web mapping applications that have a large area dedicated to the map are more usable (Nivala, Brewster, and Sarjakoski, 2008; Skarlatidou and Haklay, 2006). Thus, the map view is the central aspect of the prototype and takes by default a large of part of the display as shown in Figure 3.5. As mentioned in Section 3.3.1, the geographical data consist of two layers including a raster background image which is pulled from Google maps and a vector layer representing user-generated drawings (e.g. the VGI used in the case study is in yellow in Figure 3.5). The basic browsing of the map is done with tools present in most Web mapping applications, including zoom to full extent, zoom out, zoom in using a bounding box, pan, previous views, and next views (see Figure 3.6).

As discussed in Section 2.4 and 3.2.3, one of the issues with VGI is that geographical data can overlap and obscure other data elements when two or more people make some annotations on slightly different areas. Thus, to help identify the drawings on the map, a drawing is highlighted in blue when the mouse is paused over it. In this way, it is easier to identify the outline of each drawing even in cluttered areas. In addition, the drawings are ordered by feature type with the polygons being drawn first, followed by the lines and the points. Despite this organization some large polygons can hide other smaller polygons. Thus, the polygons are displayed according to their area with the largest ones drawn first and finishing with the smallest at the top. In this way, all polygons are at least partly visible. Finally, when clicking on one of the drawings, it is displayed on top of all the other drawings in a dark blue color. Hence, the full outline and shape of any drawing can be seen completely even if it is only partly visible in the default display.

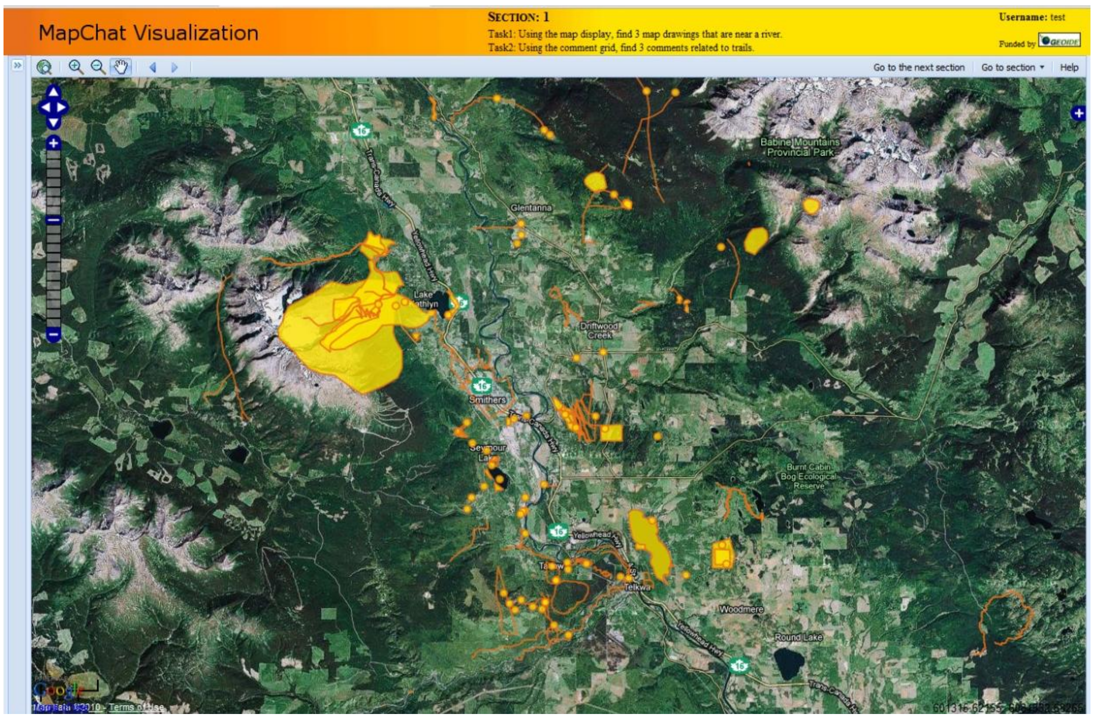


Figure 3.5: Map view

Figure 3.6: Navigation tool bar

As mentioned in the introduction of Chapter 3, the generic attribute data consists of a text note and a set of attributes. The basic browsing of the attribute data is done with a table view that displays the text notes along with the attributes (see Figure 3.7). For instance, in the case study, the text note and the user ID of the contributor were displayed as well as a contribution ID used to provide a reference to the users when the prototype is used in a testing setting with a questionnaire. In the effort to make a progressive and adaptive interface, the users can easily reveal and hide the tabular view on the left of the screen as well as adjust its size.

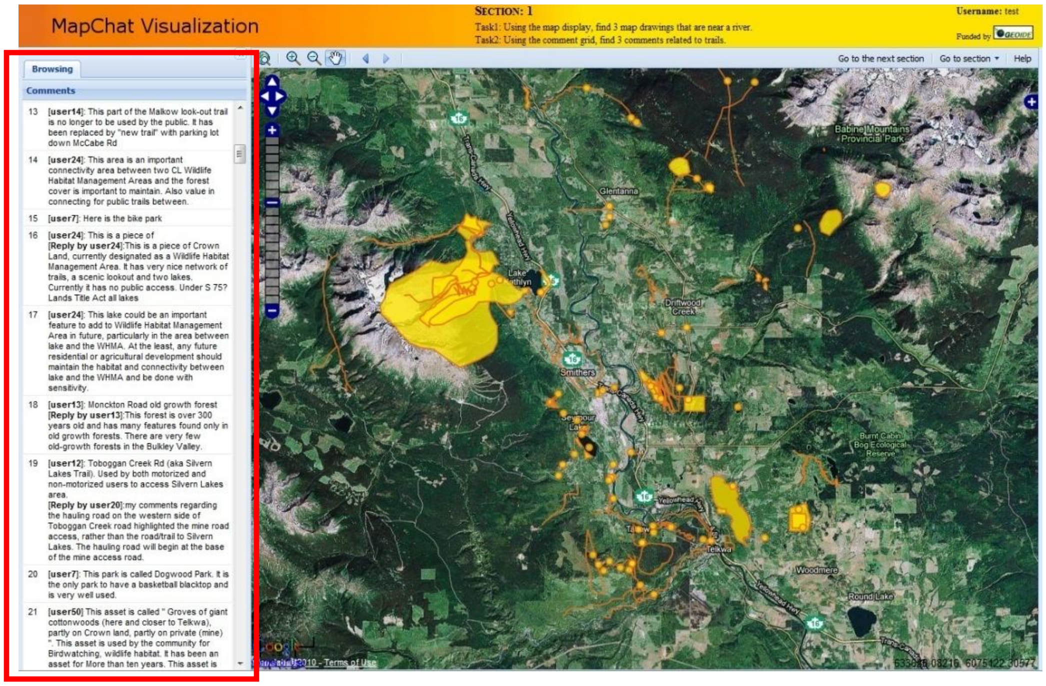


Figure 3.7: Tabular attribute view

3.3.5.2 Linkages Between Geographical and Attribute Data

In addition to browsing the attribute data and geographical data, it is essential for users to be able to reference the text notes and their linked drawings as seamlessly as possible. As discussed in Section 3.2.4, this aspect of the software is developed with techniques from coordinated multiple views research (Keim et al., 2005; Roberts, 2005, 2008; Robinson, 2009). To link the comment list to the map, the brushing and linking techniques described in Section 3.2.4 were used. When a user moves the mouse cursor over a text note, the linked drawing is automatically highlighted on the map. This allows for very quick browsing of the text notes with an instant localisation of the related drawing. To complement this feature, a "zoom to" function was implemented to allow the users to locate precisely a drawing. When clicking on a comment, the map is centered and zoomed onto the related drawing(s).

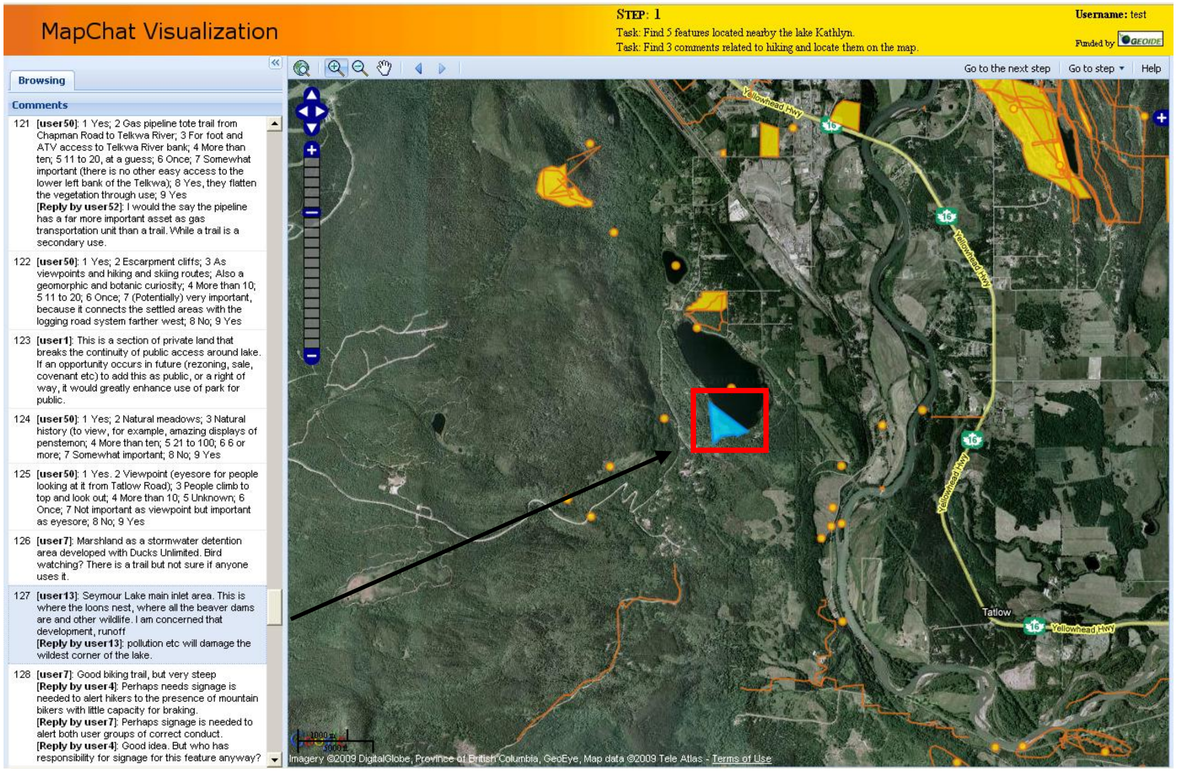


Figure 3.8: Linking and brushing

To display the text note associated with a mapped drawing, a pop-up window, also called an info window, appears when the user clicks the mouse on a drawing. The use of a pop-up window is a longstanding desktop GIS feature and it is now widespread in Web mapping applications due to its ease of use (Sheesley, 2009). In MapChat Viz, the pop-up contains the text note related to a drawing as well as its feature ID and tag features as described in Section 3.3.5.5. A brushing and linking technique could also have been implemented to establish a linkage from the drawings to the comments as explained above to make the link from the comments to the drawings. It would make the software more consistent, however, this feature was not implemented due to lack of time.

3.3.5.3 Interactive Data Filtering

As mentioned in Section 3.2.3, a simple display of the data along with basic browsing features is not sufficient to ensure an efficient data exploration and deal with information overload that occur when collecting VGI. Interactive filters allow users to constrain temporarily the data that are displayed

(Roberts, 2008). In MapChat Viz, dynamic queries are employed to allow users to filter the data display interactively based on attribute selections or geographical queries.

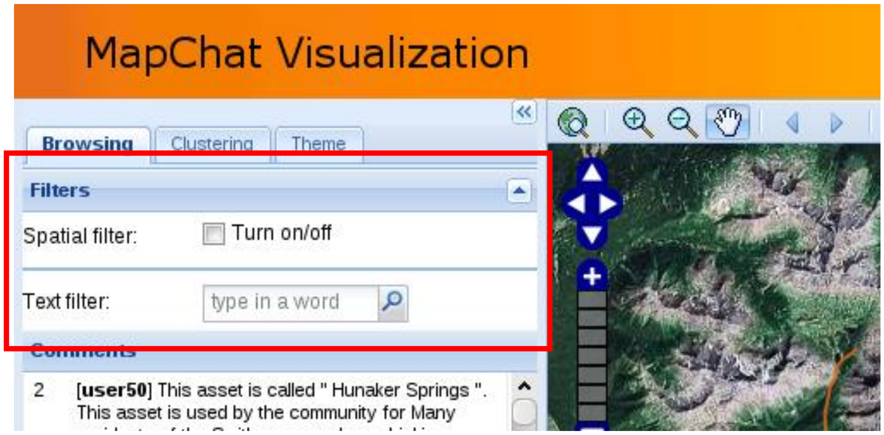


Figure 3.9: Text and Spatial filter

Attribute-based filtering can be done in MapChat Viz using a simple textbox in which users enter a string to search within the text notes. When a user enters a word in the search box, the tabular view is refreshed to display only the text notes which include this string. The search functionality does not implement more advanced features such as separate words, metaphone algorithm or logical connectors. This interactive filtering is combined with linking methods by automatically selecting and highlighting the drawings related to the filtered text notes.

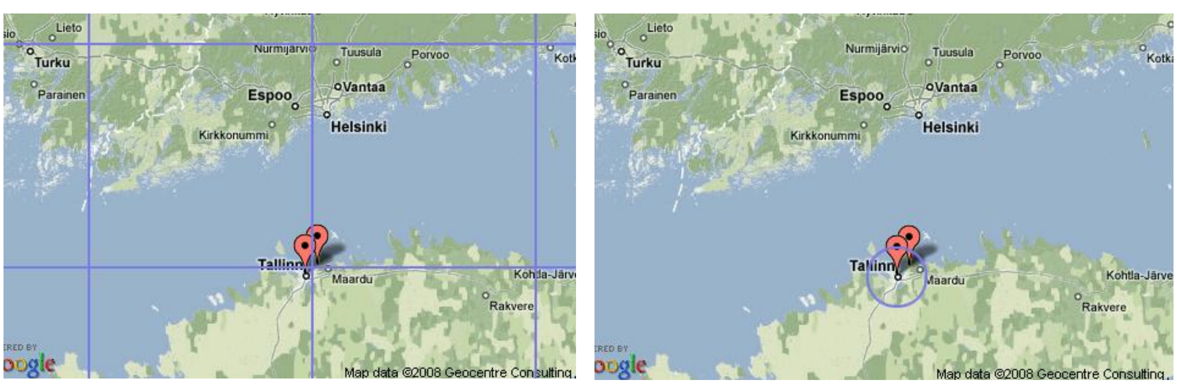
Text notes and drawings can also be filtered spatially. While spatial filtering can be performed in different ways such as a bounding box, freehand lasso, line or line brush (Roberts, 2008), only the bounding box approach is used in MapChat Viz. When the spatial filter is activated, the contribution list displays only the text notes related to the drawings contained in the area currently viewed. This method was chosen for its simplicity of use and development. Another option could be to allow the user to draw their own polygon, however this was not implemented.

In accordance with the generic needs expressed in Section 3.2.3, the two filters can be activated at the same time to work in a conjunctive way. If so, the table content is filtered according to both the spatial and the text filter. Therefore, a user is able to find all the text notes containing the word relevant to him in a given area.

3.3.5.4 Spatial Aggregation and Visualization of Spatial Repartition

As discussed in Section 3.2.3, high concentrations of VGI can occur in some areas of the map which cause the data to appear cluttered and make the data browsing difficult. Furthermore, an unequal spatial repartition of VGI can give insights on the contributed data although it has to be interpreted carefully. For instance, it can reveal areas that are of particular interest for a community but it can also reveals imbalances in the outreach and recruitment of the participants or it can simply be due to artefacts in the data such as the duplication of similar contributions. However, a spatial repartition can be hard to visualise and quantify due to visual artefacts such as clutters of features or several large imposing polygons. To ease the browsing of cluttered areas and the visualization of the spatial repartition, a dynamic data clustering feature is implemented in MapChat Viz and described below.

The first clustering methods to appear in Web mapping applications were motivated for technical reasons since map displays can be slowed significantly by the presence of numerous features. Clustering methods permit the number of features on the screen to be reduced and therefore speed up the display. The two main methods to cluster or aggregate the data are a grid-based method and a distance-based method. In the grid-based method, the map is divided into squares and if the number of features in a square exceeds a threshold the features are clustered together. This technique has several limitations. Indeed, two features can be really close but in different squares or two features can be far apart but in the same squares (see Figure 3.10). In the distance-based method, the features are clustered if their distance from each other is inferior to a clustering threshold. Considering the limitations of the grid-based method and the fact that the distance-based method is already implemented in OpenLayers, the distance based method was used.



Source: Tuupola (2008)

Figure 3.10: Comparison of the square-based and distance-based method

The clustering feature was implemented by using the programming class provided by OpenLayers. As such, the clustering was performed by looping through all the spatial features in the map (all feature types included). For each feature, the center of the feature bounding box is used to calculate the distance from the feature to each of the other features, and clusters as they get created. If a feature is within a distance threshold (chosen as a parameter) of another feature or existing cluster, a cluster is created or the feature is added to the existing cluster. The clustering is obtained once the algorithm has looped through all the features. This approach provides good performances and the clustering is performed almost instantaneously with a couple hundred features. However it has some limitations that can results in imprecise representations. A cluster is located at the center of the bounding box of the first feature that constituted it and not at the average centroid of all the aggregated features and also, the order in which the features are clustered can change from one map view to another which can consequently change the clustering representation and can be confusing for the users.

Another important aspect of the clustering method is the representation of the clusters. Their representation draws upon the classic cartographic principles developed on proportional symbol mapping (Slocum, McMaster, Kessler, and Howard, 2008b). The clusters are represented by point symbols sized accordingly to the number of features aggregated (see Figure 3.12). The size calculation follows perceptual scaling rules that take into account a correction factor for the visual underestimation of larger symbols as opposed to mathematical scaling that sizes the symbol in direct proportion to the data (Slocum et al., 2008b).

Figure 3.11: Clustering symbols

Instead of using a legend that can be hard to read, the number of features aggregated in a given symbol is directly written in its centre. In addition to the size variable, a three colour scheme can optionally be added to the cluster symbolisation (see Figure 3.11). The cluster distance varies dynamically with the scale of the map displayed. In this way, the clusters are slowly disaggregated when zooming into an area or aggregated together when zooming out. In addition, the clustering threshold distance can be adjusted by the users with a slider (see Figure 3.12) to adapt the clustering to the dataset being visualised and to study different levels of detail for a given scale. Finally, some data browsing features are adapted to the cluster representation, where a paging pop-up gives access to each comment contained in a cluster (see Figure 3.14) and brushing on a comment highlights the cluster that contains the related drawing.

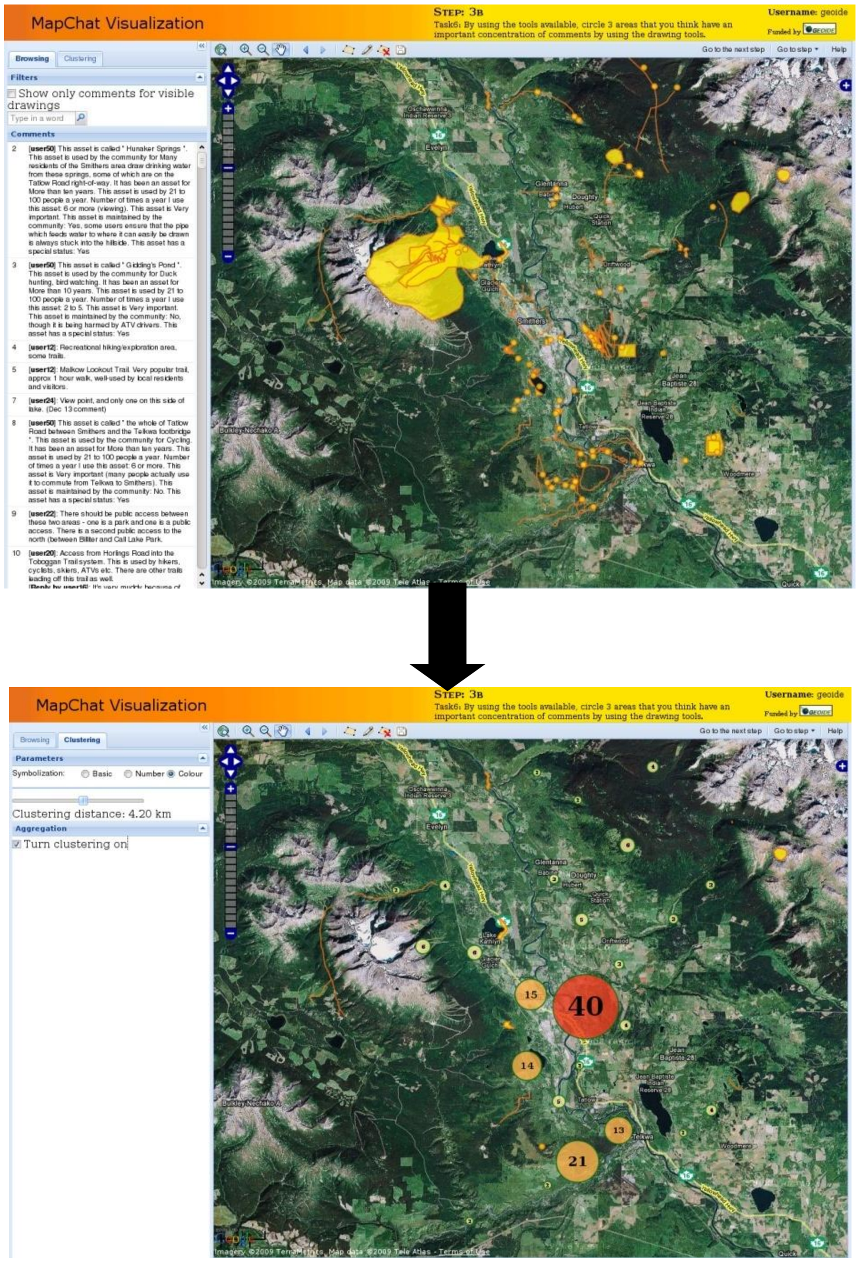


Figure 3.12: Example of data clustering in MapChat Viz

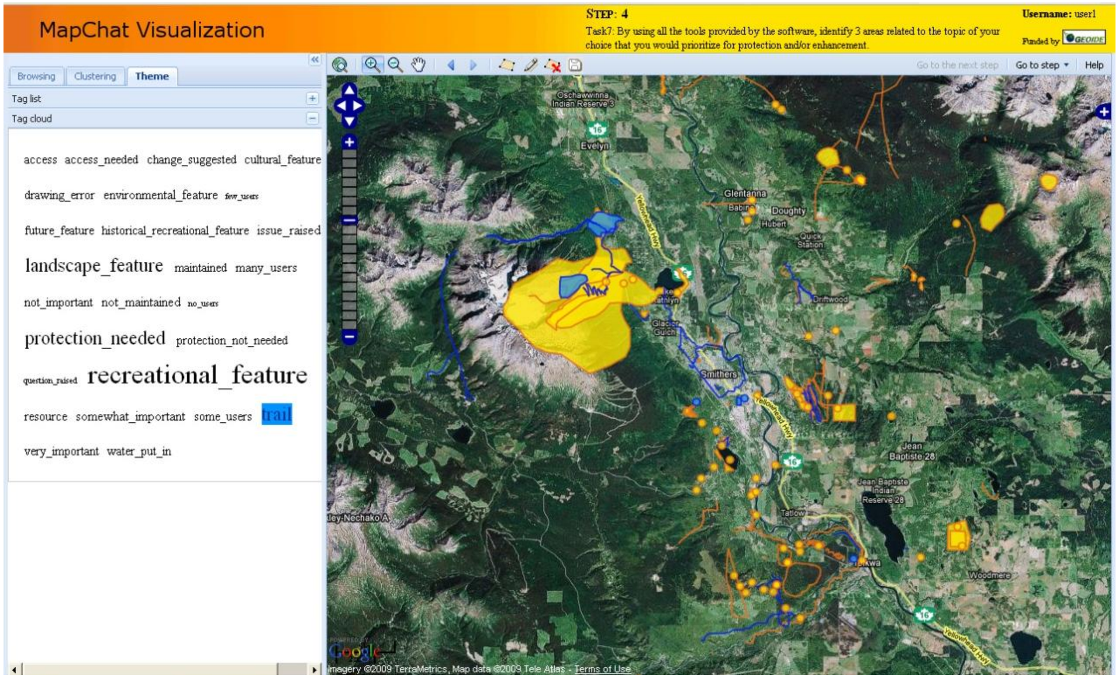
At the top the raw data are represented. At the bottom, the data are clustered in proportional point symbols sized and coloured according to the number of features aggregated

3.3.5.5 Tag-based Visualization

As discussed in Section 3.2.3, several approaches exist to categorize VGI data and it belongs to each VGI project to choose the one that is the most appropriate in their specific context. The interest in the tagging system, besides the fact that it allows data to be categorized, is how and how well it can permit the visualization of data categories in a priori non-categorized data. There are several ways to visualize tagged data (Smith, 2008). Two of them were implemented in MapChat Viz. The first is a simple tag list that displays tag names and their relative tag count. The tag names are displayed sorted in the decreasing tag count order. It provides quantitative information. The second one is a tag cloud. Tag clouds are popular on the Internet and can be seen on many blogs as a quick way to show the blog content. A tag cloud displays the tag names like a tag list, however, the tags are in alphabetical order and the tag count is represented by the font size of each tag name. It still provides qualitative information but gives more importance at the impact of the visual impression on the users.

The tag list and tag cloud were used in combination with the linking and brushing methods, and interactive filtering techniques. Indeed, the tag list and cloud are automatically refreshed after a map move to reflect the tag name and tag count contained in the area being viewed. Moreover, when a user moves the mouse over one of the tag names, all the drawings which are tagged with this tag are highlighted in blue on the map (see Figure 3.13).

MapChat Viz capitalizes on data classification and the spatial clustering to offer another method to analyse data classification. When the clustering is activated, a left click on a cluster displays, besides a pop-up, a pie chart graph representing the proportion of the number of times each tag has been applied on the data contained in the selected cluster (see Figure 3.14). In this way, the user gets an overview of the tag repartition within a cluster. Some experiments were also done with using the pie chart as an icon directly on the map. However, time constraints prevented from properly implementing it and testing it in the case study.



Note: the tag cloud is refreshed according to the spatial extent being displayed and the user is brushing the tag “trail”

Figure 3.13: Tag cloud

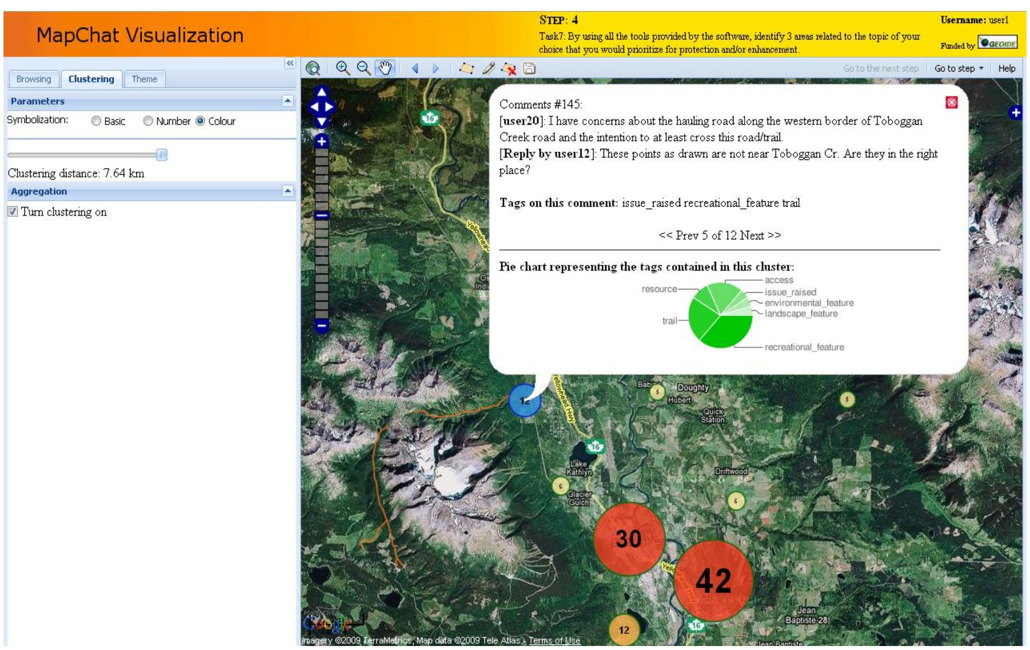


Figure 3.14: Paging pop-up and the pie chart

3.3.5.6 Evaluation Tools

MapChat Viz implements several tools to support the primary goals of experimentation and evaluation of the visualization features presented in Section 3.3.1. First, in order to be used in combination with questionnaires during workshops, the MapChat Viz interface can optionally be built in a step-wise manner by adding new pieces as the user goes through a workshop script and questionnaire. In this way, the participant attention is focused on evaluating one feature at a time which limits confounding factors and makes the results easier to interpret and compare from one user to another. Moreover, participants get to learn how to use the software progressively without being overwhelmed at the start.

The second tool implemented to use MapChat Viz as a test bed is the continuous logging of the user interactions with the software in a database to provide the researchers a source of information complementary to the questionnaire answers. The list of all the events captured along with the details stored on each action is detailed in Table 3.1.

The next section presents how this evaluation tools were used in a case study.

Events Recorded	Data Stored
Section change	Next (to go the following section) or Go to (to go directly to a section)
Tab change	The name of the tab becoming active: browsing, aggregation or theme
Map moves	The zoom level and the map extent at the end of each map moves (pan or zoom)
West Panel opening or closing	Expand or Collapse
Pop up opening	Feature Id
Zoom to from a comment to a feature	Feature Id
Mouse over a comment in the grid	Feature Id
Text filter	The keyword used for the search
Spatial filter activation	Activation or deactivation
Drawing	Zoom level and Geometry Id
Clustering	Activation or deactivation
Moving the clustering slider	The new clustering distance
Cluster symbology change	Symbology name: basic, colour or number
Paging through the pop-up	Next or previous page
Tag cloud display	Collapse or expand
Tag list display	Collapse or expand
Mouse over a tag	Tag name

Table 3.1: Logged events and the related stored data

3.4 Chapter summary

Chapter 3 has presented the design of the MapChat Viz prototype software. First, the generic needs and general concepts underlying the software development were examined. These needs and concepts guided the practical development of the prototype software was presented, including the technical choices that were made, the general software architecture and the implementation of the visualization tools. The next chapter describes the design of the case study that was carried out in order to evaluate of the prototype application.

Chapter 4
Case Study and Research Design

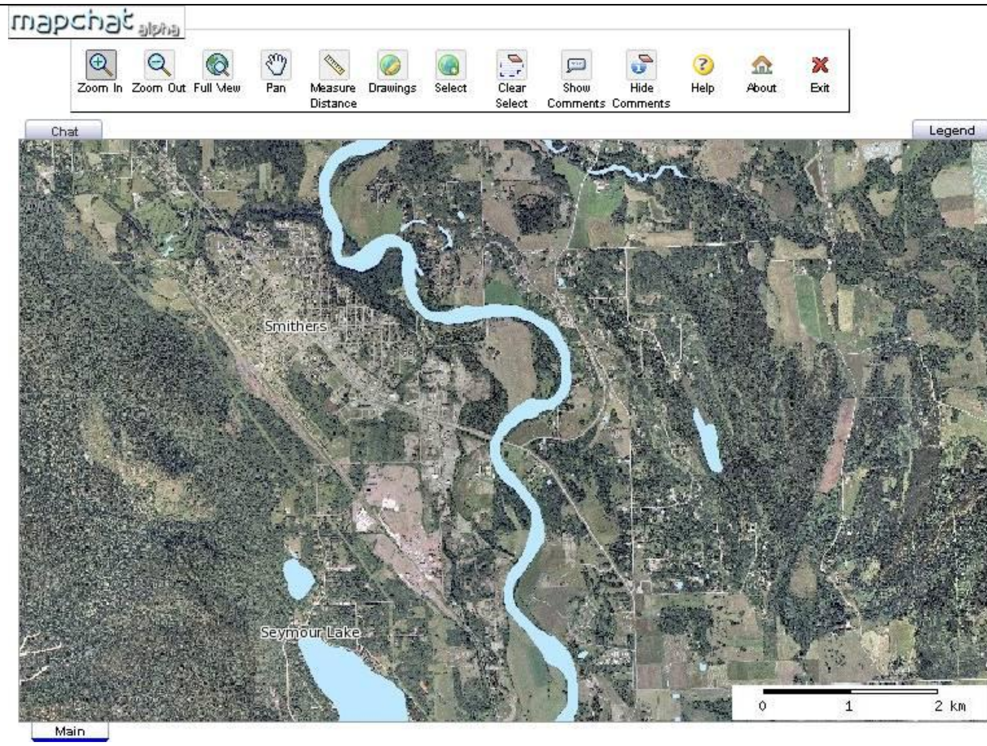
This chapter presents the design of the case study in which the prototype was evaluated. Sections 4.1 to 4.3 provide background information on the study area, the Bulkley Valley in British Columbia, Canada and present how the MapChat software, developed by a previous research project, was used to collect a set of Type 2 VGI that was used to evaluate the prototype functionality. Section 4.4 discusses the research methods underlying the collection of data on the use and evaluation of the visualization tools by the participants. Finally, Section 4.5 describes how these methods were used in a series of workshop to evaluate the prototype application and the techniques and concepts underlying its design.

4.1 MapChat

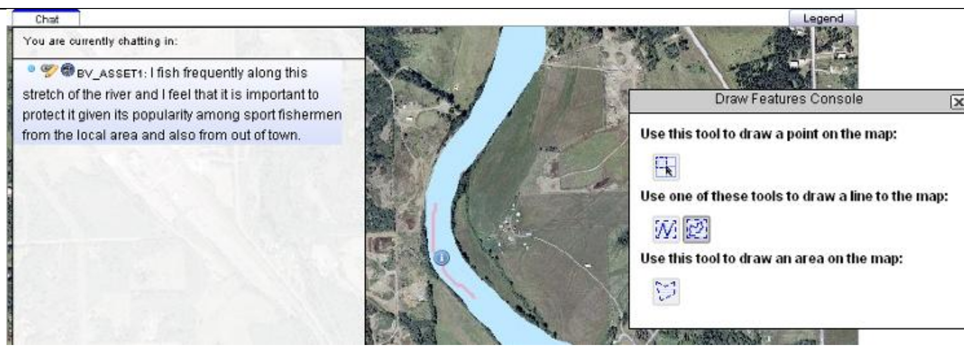
4.1.1 Overview of MapChat

The MapChat Viz study belongs to a broader project entitled "Promoting sustainable communities through participatory spatial decision support" funded by Geoide (GEOmatics for Informed Decisions). The overall goal of this project was to develop and evaluate computer-based mapping tools designed to facilitate collaboration and decision-making among stakeholders in community planning processes. One of the main pieces of software developed to meet this goal is called MapChat, which was briefly presented in Section 2.3.4.2. MapChat is a Web-based application designed to facilitate map-based discussion between different stakeholders such as local people, planners and policy makers or anyone who has an interested in the issues being discussed. Practically, users can draw features (point, line, and polygon) or select existing features on a map. Then, they can leave comments linked to specific features or drawings. Besides drawing their own features and commenting on them, the users can also interact with the other users synchronously or asynchronously and reply to others' comments to agree, disagree or simply add details on what has been previously said. Therefore, a map-based discussion with multiple threads takes place as illustrated in Figure 4.1.

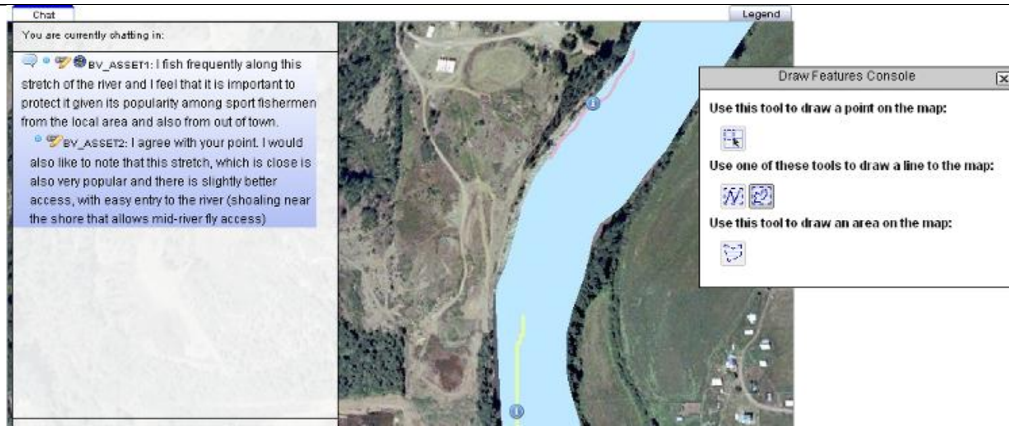
This process results in the creation of a rich Type 2 VGI dataset with data that are qualitative, subjective and heterogeneous as illustrated by the data collected for the case study described in Section 4.3.



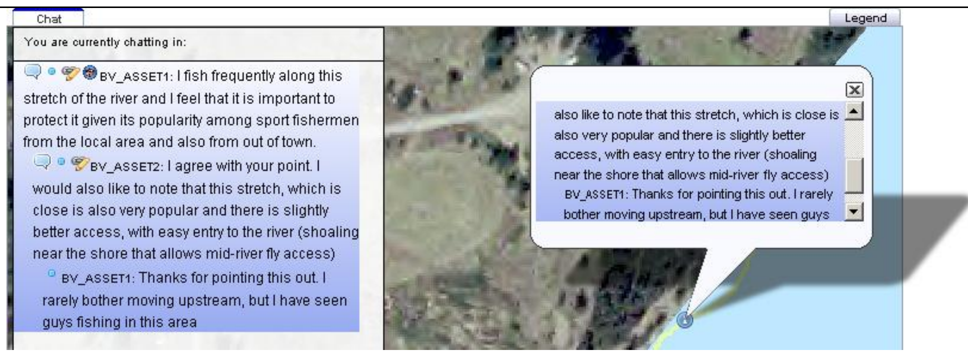
User1: Opens MapChat and uses the Zoom Tool to zoom to a stretch of the Bulkley River



User1: Draws a line along a stretch of the river and comments on the asset value of this location to fishermen. The pencil icon to the left of the message shows that the comment is linked to the feature in the map view.



User2: Replies to User1, extending the thread with a response after zooming to the location of User1's comment. User2 adds a further stretch of the river upstream from User1 and links a new comment to this new (extended) asset location.



User1: Responds to User2 by adding to the thread with a linked comment, but no additional asset location is mapped. The comment bubble in the map view displays all comments linked to an asset.

Source: Hall et al. (2010)

Figure 4.1: MapChat user interface and discussion sequence between two participants

MapChat draws upon the advantages of Web-based PPGIS to enable users to take part in discussions and to make the participation process more accessible. It has been used in previous case studies dealing with issues including collaborative assessment of affordable housing in the town of Collingwood, Ontario (Noble, 2007; Taranu, 2009). MapChat version 1 was built with open source software including PostGIS, MapServer and Chameleon. A full description of the system is available in Hall and Leahy (2008). Version 1 of MapChat used in this thesis has been since superseded by an updated version 2 described in Hall and Leahy (2010) that can be found at http:// mapchat.ca.

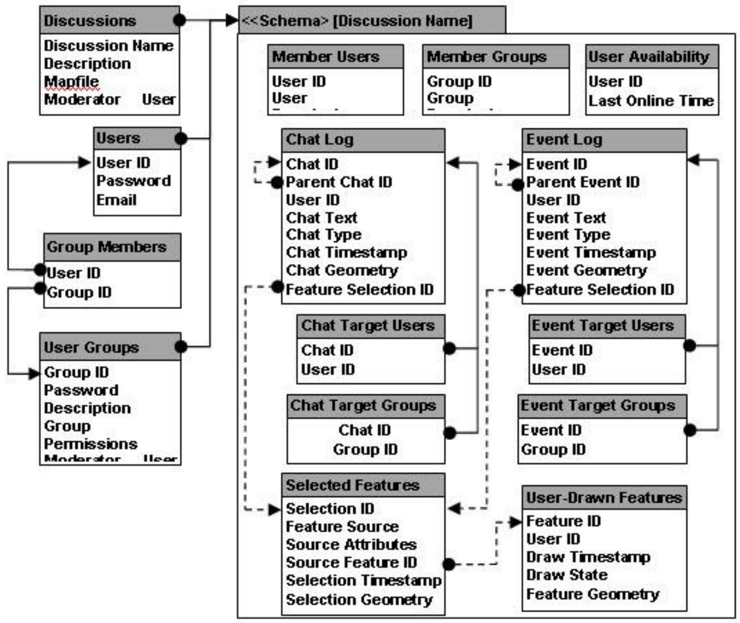
As further detailed in the next sections, the user-generated chat and spatial data that can be collected with MapChat are a good example of Type 2 VGI. Thus, MapChat was used in this study to collect a VGI dataset representative of Type 2 VGI that was then loaded and formatted to be used with MapChat Viz since it does not integrate features to collect map-based discussion but focuses on visualization and browsing. The next sections describe the generic data structure used by MapChat and how it was adapted to be used into MapChat Viz before presenting how MapChat was used in a case study to collect a VGI dataset.

4.1.2 MapChat Data Structure

MapChat stores two kinds of data, data that are related to the user management system (e.g. users, groups of users, different level of rights) and the data that users generate using the software. The former data were not used in the case study as MapChat Viz focuses on the visualization of the user-generated data and does not incorporate the contribution process.

The user-generated data collected with MapChat is organized in a set of discussions that are created by the administrators of the software and relate to a geographic area under examination. Typically, a discussion is focused on a specific topic or issue. For each discussion created, a schema is generated in the database with the structure illustrated in Figure 4.2. A discussion contains two main types of data records namely user drawings and comments. The comments are recorded in the "Chat log" table along with the time, the user and a reference to the parent comment if the comment is a reply. The reference to the parent allows for the reconstruction of the threaded discussion with a recursive query. User drawings are stored in the "User-drawn feature" table with the geometry stored as a spatial object. MapChat allows users to link a drawing to any number of comments and also to link their comment to any number of drawings. Therefore, there is a m-m relationship between the comments and the drawings. To enable this m-m relationship, the relationship table "Selected feature" is required.

Overall, the structure of MapChat data constitutes one example of the generic Type 2 VGI as defined in Chapter 2 and used in Chapter 3 discussion and is thus appropriate to be used with MapChat Viz in a case study. However, MapChat Viz uses a simpler generic data structure so the data collected with MapChat need to be processed and reformatted before being used as described in the next section.



Source: Hall and Leahy (2008)

Figure 4.2: Database schema for MapChat discussions

4.1.3 Data Conversion and New Data Structure

The user-generated data collected with MapChat need to be cleaned and processed to be used with MapChat Viz. The main data conversion and cleaning procedures are summarized below and further detailed in the reminder of this section:

1. Delete erroneous comments
2. Make comments anonymous by removing information that identifies individuals
3. Concatenation of the initial comment and the replies linked to each drawing
4. Merging of the drawings that have the same sequence of comments
5. Aggregation of the discussions into one discussion
6. Simplification of the spatial geometries.

The first version of MapChat was intentionally designed to prevent users from deleting their comments in order to mimic real discussions where what had been said cannot be deleted. However, some comments were clearly errors related to users' unfamiliarity with the software. For the purpose of this study, these error comments were excluded. They were identifiable by the fact that they were usually followed by a message such as "Delete this comment". In addition, any references to individuals' name within the comments were deleted and replaced by anonymous names to protect privacy.

As explained earlier, there is an m-m relationship between drawings and comments in MapChat. Although it reduces the accuracy of the information in a few cases, the relationships were simplified to 1-1 relationships to ease the development of the prototype as well as to make the data presentation easier for users to understand. To obtain a 1-1 relationship, all comments linked to a drawing were concatenated together. As a result, the dataset contains a set of individual drawings with each drawing linked to one concatenated comment that itself may contain one or more comments. Hence, at this stage, there are different drawings linked to the same concatenated comments (or sequence of comments) (see Figure 4.3). For visualization purposes, and especially for highlighting drawings, the drawings linked to the same sequence of concatenated comments were merged together in a single spatial feature. Once each discussion was processed according to these steps, they were merged together into a single dataset.

Finally, after merging all of the discussions together, the number of geometries was large (about 250) and this started to slow down the application because of the vector format necessary for visualization that is heavy to handle for the Web browser. Thus, in order to improve the speed of the application, the geometries were generalised by using the PostGIS simplify function. This function uses the Douglas-Peucker generalisation algorithm (Slocum et al., 2008a) to reduce the number of vertices and therefore the complexities of the geometries. The simplification of the geometries was performed with a small tolerance so that it does not adversely adjust the shape of the information but rather just corrected some of the very small scribbles made in error by some users when using the drawing tool. The generalisation of the geometries made a significant improvement in the performance of the application.

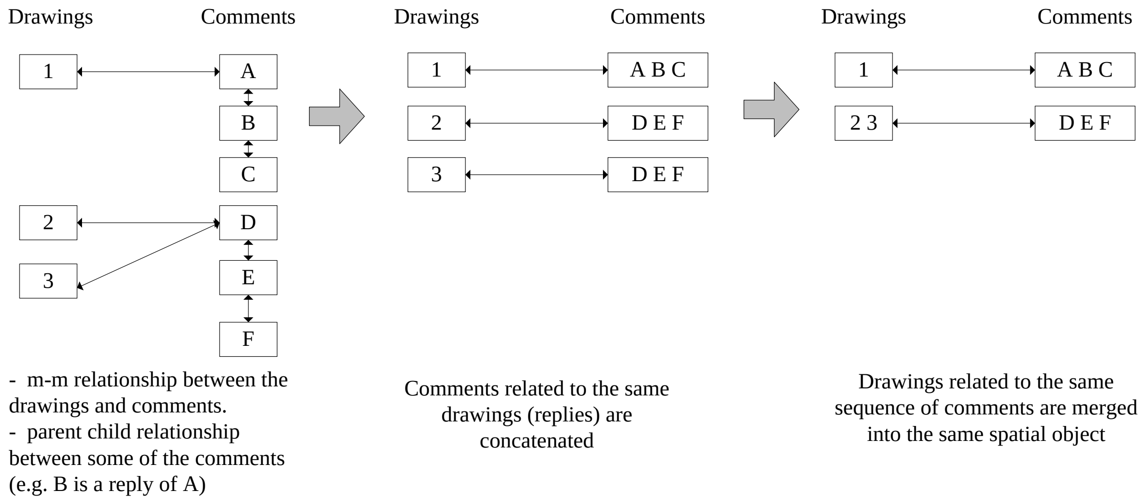
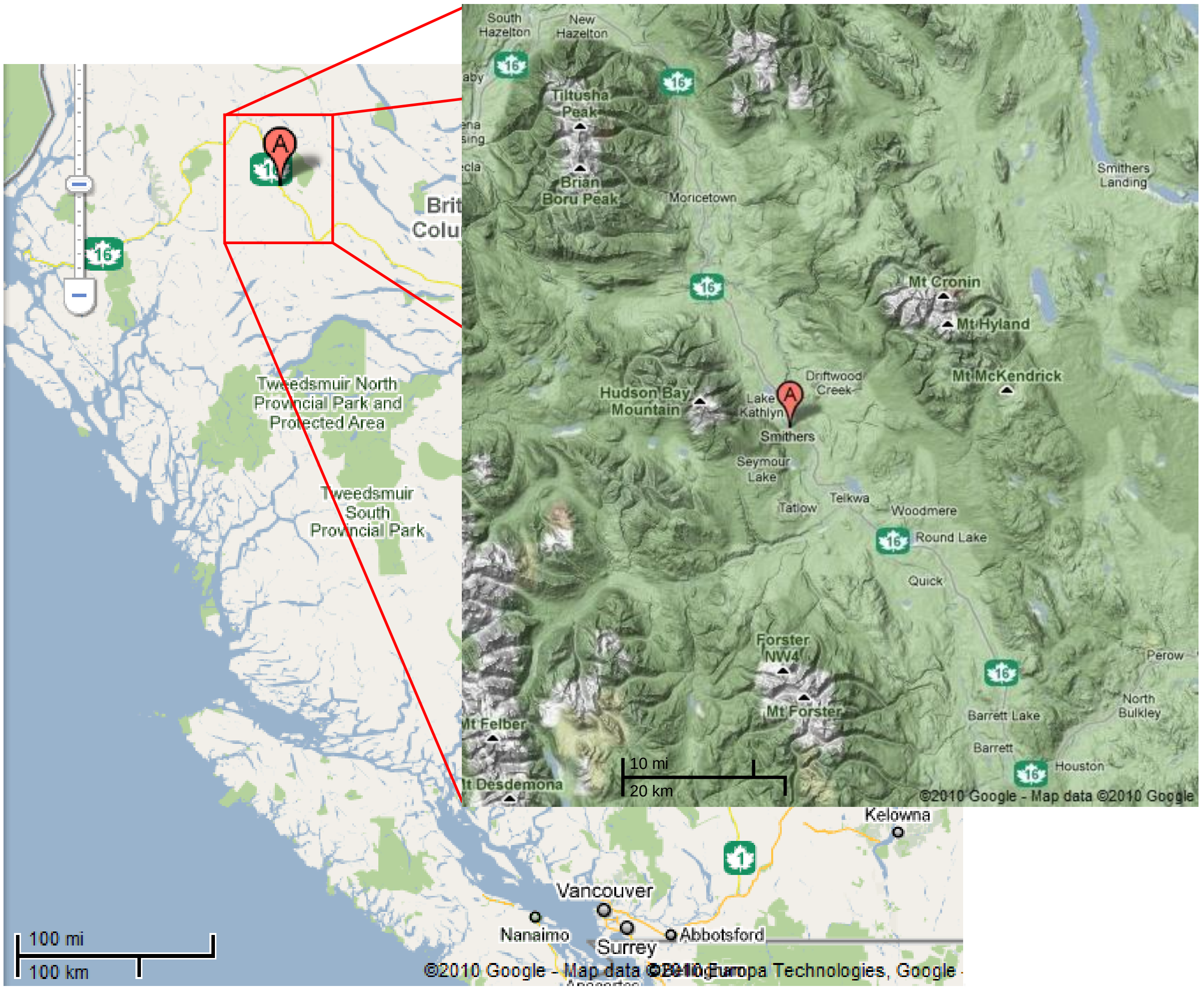


Figure 4.3: Illustration of Steps 3 and 4 of the data conversion process.

4.2 Study Area

Due to the fact that one of the researchers resided there, and the particular characteristics of that region, the Bulkley Valley was chosen as one of the study areas for the project "Promoting sustainable communities through participatory spatial decision support" (see Section 4.1) including the case study presented in this thesis. The Bulkley Valley is located in north western British Columbia in a relatively remote location about 1000 km north of Vancouver. It is an open plain oriented south east to north west delimited by the coastal mountain range on the west side and several mountain ranges on the east side (see Figure 4.4).

The surrounding mountains provide natural resources and opportunities for outdoor recreational pursuits, while the valley floor offers opportunities for agriculture, forestry and small urban settlements. The overall population of the Bulkley Valley is approximately 20,000 people distributed amongst several communities, which include from the south end of the valley to the north Houston (pop. approx. 3,500), Telkwa (pop. approx. 2,000), Smithers (pop. approx. 5,400), and the Hazeltons (pop. approx. 3,000). These communities include territories of the Wet'suwet'en First Nation which actually encompass most of the Bulkley Valley and extend even further in some parts.



Source: Google maps (see details on the maps)

Figure 4.4: Location of the Bulkley Valley

Most of the Bulkley Valley land is owned by the Crown (under the auspices of the Federal and Provincial governments in Canada) and has been under development pressure over the last decade for ventures including the expansion of a ski hill, proposals for new mines, new rural subdivisions and timber harvesting. These potential developments endanger the valley amenities that are not formally inventoried and mapped. Even though most residents of the Bulkley Valley have developed a strong respect and responsibility for their local environment, it is difficult for them to contribute their local knowledge to the planning process in order to protect amenities that are unknown to local, regional and provincial planners and policy makers (Chipeniuk, 2004, 2008). Generally, opportunities for public inputs occur too late in the planning process (e.g. when proposals are already threatening some of the valley amenities) to ensure a proper assessment of the potential impacts of the development plans on valued community assets.

To address these issues, various local groups formed to foster a proactive and community-led planning approach. This context provides an excellent opportunity to leverage the power of PPGIS as a community-based and participatory approach to inventory and discuss local knowledge before developments occur. The specific approach that was used in the Bulkley Valley consisted of inventorying, mapping and discussing the community's most valuable social, natural and cultural assets. Such an approach is referred to as 'asset mapping' in the asset-based community development field (ABCD) (Fuller, Guy, and Pletsch, 2002) or as cultural mapping within the cultural planning realm (Baeker, 2010). In this project, a community asset can be broadly defined as "any natural, built or cultural feature that a citizen can map and may deem worthy of sustaining. Assets can be a point in geographic space (e.g. a lookout), a linear feature (e.g. a stretch of river), or an area (e.g. a lake). The central underlying commonality is that assets are based upon citizens' local knowledge of the community they live in" (Hall et al., 2010, p. 7).

4.3 Community Asset Workshops and Community Asset Data

4.3.1 Community Asset Identification Workshops

A research team composed of several researchers and students involved with MapChat research organised a series of workshops in the Bulkley Valley in partnership with two local organisations, namely the local office of the British Columbia Ministry of Environment (MOE) and the Bulkley Valley Stewardship Coalition (BVSC). These partnerships were set according to the principles of Community Based Participatory Research (CBPR). A CPBR approach sought to make the most of the researchers' knowledge and citizens' local knowledge by enabling co-production of knowledge and mutual learning. This allows the researchers to consult local people regarding their needs and identify the most relevant procedures to make sure that the study is relevant in the local context. Also, feedback on the research tools and methods can be collected. Local communities are empowered by access to state-of-the-art technology that provides them with access to their local information while facilitating the collection of their local spatial knowledge. Moreover, researchers can help build local capacity in the use of the software which ensures that the community will benefit from the study in the mid and long term and that their own knowledge is disseminated and integrated with the other knowledge bases that are typically drawn upon in decision planning. This includes spatial data layers collected by official agencies, such as conservation and habitat zones defined by the scientific community.

The Bulkley Valley workshops focused on several topics including riparian zones at risk, a trail inventory and community assets. For the study presented in this thesis, only the data collected on community assets were considered. The series of community asset identification workshops was organized in two phases. First a series of three, four-hour workshops was organized jointly by the university-based researchers and the BVSC between October 2008 and January 2009. These first three workshops were held in a small group context in a computer lab at the Northwest Community College in the town of Smithers. The researchers first made a quick presentation of the purposes and uses of the MapChat tool and then assisted participants with the use of MapChat for asset identification. In this first phase, local people were able to learn how to use MapChat software and to start contributing their local knowledge with it. By building their skills in this face-to-face context, participants were able to connect subsequently to the MapChat tool from their home over the Internet and further contribute asset data at times convenient to them. This first series of workshops also allowed local organizations to build capacity to initiate and organize a new series of self organized workshops without the help of the researchers. The organization of these later community-led workshops was divided up by neighbourhood where participants from specific communities gathered in a series of informal meetings held at individuals' homes.

4.3.2 Community Asset Data

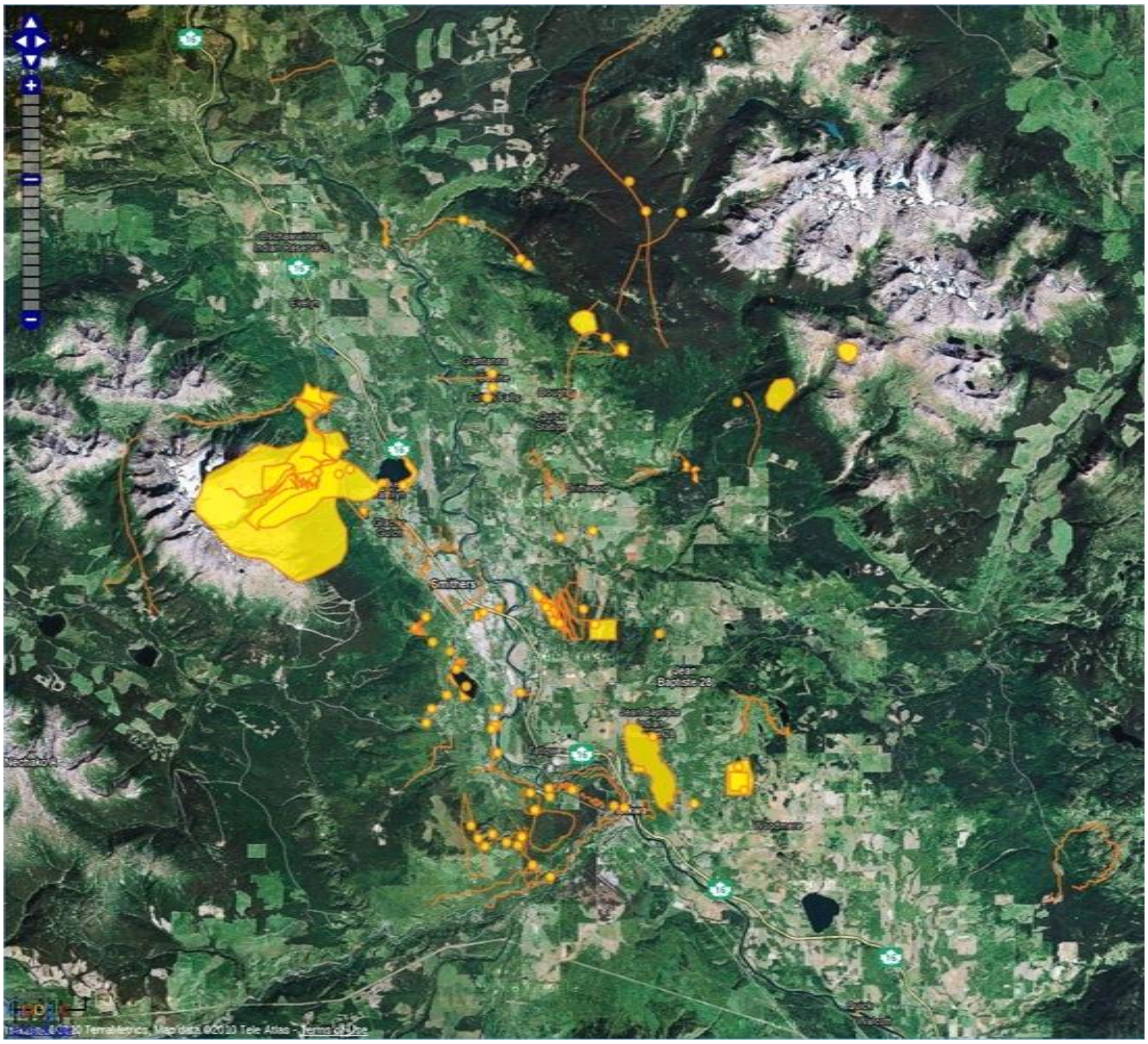
The asset data that were compiled and used in the case study comprise outputs generated during the MapChat workshops organized by the researchers as well as the data collected during the community-led neighbourhood workshops from their initiation to the time of writing. The neighbourhood workshops are still ongoing in the Bulkley Valley but only the first three of these are included in this study. Table 4.1 presents basic metrics of the asset data used in this thesis to evaluate the merits of MapChat Viz. The first three columns summarize the raw data collected in MapChat. The fourth column presents the number of drawings and concatenated chat threads after these were merged according to the process presented in Section 4.1.3. Finally, the last column shows the number of records that were used in the MapChat Viz case study after the removal of obvious erroneous comments. Hence, the final number of records used is two hundred. It is not possible to state whether this is a large or small number in the absence of similar benchmark. However, more information and analyses on the asset data can be found in Hall et al. (2010).

Discussion	Total messages	Total threads	Drawings	Drawing and concatenated threads after merging	Drawings and concatenated threads after cleaning error comments
Research workshop1	30	22	41	20	16
Research workshop2	56	46	61	45	40
Research workshop3	22	15	25	14	11
Neighbourhood workshop1	53	50	80	45	36
Neighbourhood workshop2	71	63	82	54	51
Neighbourhood workshop3	71	56	61	54	46
Total	**303**	**252**	**350**	**232**	**200**

Table 4.1: Amounts of the asset data collected during the Bulkley Valley MapChat workshops

The table presents the amount data before and after being processed to be integrated in MapChat Viz accordingly to the procedure described in Section 4.1.3. In dark grey is the final number that was used for the case study.

This results show that MapChat has been successful at fostering community involvement through citizen-based data collection (Hall et al., 2010). As the success grows, some of the VGI challenges described in Section 2.4 and 3.2 are becoming more and more apparent. Indeed, as the number of users and contributions is growing, the data are becoming significantly harder to browse and understand to the point that the utility of the data is significantly reduced. The number of contributions makes manual browsing tedious and some areas of the map get an accumulation of drawings that overlap each other making their visualization impossible (see Figure 4.5).



Note: the user-generated asset data is in orange

Figure 4.5: Asset data collected during the Bulkley Valley MapChat workshops

The browsing of the data in MapChat is also made difficult by the unstructured nature of the chat comments. Unfortunately, MapChat does not support tag creation so this process could not be evaluated and experimented with. However, in order to evaluate the merits of such data classification methods and related visualization tools, a post hoc approach to tagging of the data was employed by the researchers. Tags were applied by the researchers in the folksonomy tradition. One or more tags were applied to each feature based on the type of asset (see Table 4.2). The tags were freely chosen by the researchers depending on the content. Another solution could have been to follow a preestablished taxonomy structure such as the cultural resource framework defined for cultural planning practices by Baeker (2010).

A single map feature can have several tags because it can be multi-functional in nature (e.g. a park area that is used for environmental and recreational purposes) or because different community members may have added text comments that suggest that the feature has different meanings to them (e.g. a spot used for fruit picking by some and biking by others).

Tags	# of occurrences	Tags	# of occurrences
recreational feature	83	cultural feature	15
trail	68	future feature	12
landscape feature	64	water feature	7
protection needed	47	resource	6
environmental feature	36	access needed	6
access	33	historical recreational feature	4
issue raised	24		

Table 4.2: Summary of the tags used to characterize the features

Overall, the community data collected with MapChat in the Bulkley Valley is symptomatic of several of Type 2 VGI challenges discussed earlier in this thesis. It is thus a relevant and appropriate dataset to use with MapChat Viz for the evaluation of visualization tools in a series of workshops as described in the next section.

4.4 Human-computer Interaction, Usability Engineering, and Geovisualization

4.4.1 Goal and Methodological Approach

This case study seeks to determine the extent to which the features implemented in the MapChat Viz geovisualization tool are usable and useful for community-based participants. Do the tools enhance the exploration of the VGI collected within MapChat? Were the users able to learn how to use the tool efficiently? Did they enjoy using the tool? Beyond the assessment of the particular tools in MapChat Viz, the primary goal is to derive general conclusions about people's understanding of the concepts and methods employed and to improve our understanding of VGI.

To address these objectives and research questions properly, the user group of the tool had first to be identified (Preece, 1993). Users of PPGIS and in, a broader context, people contributing and using

VGI on the Internet, can by nature range from casual users to experts. It is therefore a very heterogeneous group with different levels and types of computer literacy, knowledge and cultural backgrounds. In this context, adopting a user-centered design approach including methods from human-computer interaction (HCI) and usability engineering proved to be useful to investigate how the features are understood and used to improve the design and usability of the tools themselves (Haklay and Tobón, 2003; Sidlar and Rinner, 2007). However, while many previous studies have focused on pure PPGIS applications, MapChat Viz also integrates aspects from geovisualization and data exploration in its design and implementation. In this context, it has been shown that HCI and usability methods have limitations (Fuhrmann et al., 2005; Slocum et al., 2001; Tobón, 2005). This study also had resource and logistical constraints that needed to be taken into account when choosing the usability methods to use. Thus after reviewing fundamental HCI and usability engineering principles in the following section, these methods are then discussed in the context of geovisualization. Finally, the last section presents the methods chosen for this study along with their advantages and disadvantages.

4.4.2 Human-computer Interaction and Usability Engineering

Human-Computer Interaction (HCI) is defined as follow by Preece et al. (1994, p. 26):

> "HCI is concerned with understanding, designing, evaluating and implementing interactive computing system for human use."

By understanding the factors involved in human-computer interaction, the HCI field seeks to create tools and techniques which help programmers to design computer systems that are efficient while being easy and enjoyable to use. It is an interdisciplinary field that includes knowledge and practice from fields such as computer science, cognitive psychology, social and organizational psychology, ergonomics and engineering (Preece et al., 1994). Usability engineering is one of the key concepts of HCI. It focuses on the evaluation of how well a person can use a computer system. Fuhrmann et al. (2005, p. 554) provide a definition adapted from the International Organization for Standardization (ISO):

> "the extent to which a system can be used by specified users to achieve specified goals with effectiveness [the extent to which a goal is reached], efficiency [the effort to reach goals], and satisfaction [the user's opinion on system performance] in a specified context of use"

Nielsen (1993) identifies five components of usability, namely learnability, efficiency, memorability, errors and satisfaction. Learnability reflects how easy it is to understand and learn the system. Efficiency is the level of productivity a user can reach once he/she has learned the system. Memorability is how easy it is to use the system again after a period of not using it. Errors are the times users spend on performing actions that do not result with the expected outcome. Satisfaction is the pleasure and the satisfaction that a user gets when using the software. As presented in the previous section, this study focuses mainly on the learnability and user satisfaction parameters defined by Nielsen.

The different approaches to software usability studies can be classified in several ways. One way is to use the type and level of control that a researcher has on the study (Kirakowski and Corbett, 1990). The first type is a naturalistic study which takes place in a real world context where investigators play only a background role. Quasi-naturalistic studies also use a real world context, but researchers have some control on the study to collect information. Finally, experimental studies are designed to isolate some independent aspects that the researchers want to study and avoid confounding elements. This classification is comparable to one proposed by Preece (1993) that includes analytic, expert, observational, survey and experimental. Another categorization, proposed by Nielsen (1993), is to use the purpose of the study. This way, a summative evaluation aims at evaluating the quality of an interface, for instance to compare two software packages and a formative evaluation aims at getting feedback from the users to improve an interface as part an iterative design process. As detailed further in Section 4.4.4, this study was designed as an experimental and formative study.

One of the fundamental usability engineering techniques is to analyse and segment the prototype functions of a software product into a set of tasks that are performed by the test users. The evaluation of the users' ability to perform each task is then used to assess the usability of the features. There are various methods and tools available to collect data on user performance. Such data can be objective or subjective and allow for qualitative and quantitative analysis. For instance, a study can involve methods to measure quantitatively users' performance such as recording the time taken to complete a task or logging of the interactions with the software (tracking and counting user actions). On the other hand, qualitative feedback from the users can be obtained through interviews or questionnaires. A more comprehensive list of the available methods and their relative advantages is provided by Nielsen (1993), Preece (1993), and Preece et al. (1994). The approach used in this study is described further in Section 4.4.4.

4.4.3 Limitations in the Evaluation of Geovisualization Tools

Usability evaluations are designed to assess how well a user can perform a sequence of defined tasks to achieve a goal with a computer system. However, exploratory and geovisualization tools are designed to support knowledge discovery processes to solve ill-defined problems where the goals might be unknown at the outset. Thus, it is a complex issue to define precise tasks and challenging to apply usability principles in practice (Andrienko et al., 2002; Fuhrmann et al., 2005; Slocum et al., 2001; Tobón, 2005). For instance, quantitative performance measurements based on the time to achieve specified tasks is not always relevant. Indeed, a test user could take great care in the discovery process over a long period of time, while another person could complete the task without much thought or effort (Tobón, 2005). On the other hand, since goal achievement is hard to measure, additional information is required to evaluate geovisualization tools, especially with respect to information about how useful individuals believe the tool to be. Indeed, usability testing methods do not really distinguish between usable and useful, therefore qualitative information about the usefulness of the tool in support of the discovery process is needed for a complete evaluation (Fuhrmann et al., 2005).

An additional difficulty when applying usability methods to geovisualization tools is that it is complex to determine if the evaluation outcomes are related to the specific software implementation or to the concepts underlying it (Andrienko et al., 2002; Fuhrmann et al., 2005). Nevertheless, as mentioned in Section 4.4.1 and pointed out by Andrienko et al. (2002, p. 327), the aim of the case study in this thesis and of geovisualization researchers overall is "to evaluate a certain technique in general, i.e., as a concept, irrespective of a particular implementation" rather than the software itself. However, in order for it to be evaluated, a concept needs first to be implemented in a specific piece of software and thus the software and the concept are tested rather than only the concept itself. Thus, deriving information about the concepts from test results must be done carefully as the specificities of the implementation, such as the user interface or bugs, can have a strong impact on the results. As a general guideline, positive results can be interpreted in favour of both the concept and its implementation, whereas negative results do necessarily not mean the failure of the concept but it may suggest that the implementation should be revised (Andrienko et al., 2002).

As a result of these limitations, it is essential to consider a combination of methods including techniques tailored with an emphasis on qualitative feedback from the users.

4.4.4 Methods Used and Data Collection

Given the usability engineering principles discussed above and their limitations for evaluating the geovisualization concepts, a set of methods was chosen and applied in this case study. These choices, described in the following section, were guided by the need to collect different types of data to support the evaluation of the features implemented in MapChat Viz as well as by practical constraints of such as the availability of staff, equipment, participants, and time.

To limit the confounding elements and facilitate the interpretation of results, the study was performed in an experimental setting. Specifically, the study was divided into separate sections each focusing on the various features represented in MapChat Viz. In each of these sections, participants were asked to perform a series of tasks defined in accordance with the usability engineering principles discussed before. Numerous techniques are available to define the tasks, including hierarchical task analysis, cognitive task analysis, or using a list of the intended uses for the product (Preece, 1993). In this case, the tasks were defined according to the features implemented in MapChat Viz in relation to the issues defined from the literature review (see Section 2.4 and 3.2) on VGI application and the observation and feedback from participants in the MapChat 1 workshops. Putting the users at the centre of the task definition is especially important when trying to define tasks for exploratory tools (Fuhrmann et al., 2005).

One of the main usability methods discussed in the literature is the observation of the users while they are using the software (Nielsen, 1993; Preece et al., 1994). This observation can be direct, where the observer visits some users and observes them using the software, or indirect, which involves audio and video recording. Indirect observation is usually preferred as it is unobtrusive. However, it requires a purpose built laboratory or at least important resources that were not available for this thesis (such as recording equipment). Thus, direct observation of users was employed to gather qualitative information. A disadvantage of direct observation is that it may influence participant behaviour. This phenomenon is known as the Hawthorne effect (Preece et al., 1994). However, an important advantage of direct observation is that it permits developers/researchers to discover unexpected uses of the software (Nielsen, 1993). This is especially important in the case of exploratory tools like MapChat Viz, as the definition of the tasks is difficult and may need to be refined (Tobón, 2005). To perform direct observation, the observer should be quiet and provide very little help to limit influencing the subjects' thought processes. However, to maximize the feedback from users and limit the time spent on the workshop, the facilitator interacted with the test users occasionally, especially when the user encountered an issue that prevent their progress. In this case, the facilitator used the simplified “thinking aloud” method (Nielsen, 1994). The users were asked to formulate their issues verbally while using the system. This provided in-depth, subjective and qualitative feedback from the users. However, human resources were limited to one facilitator therefore observations and thinking aloud methods could not be conducted systematically with every user but rather as a complementary method while conducting the workshop.

Two methods were applied to collect data consistently. A questionnaire was integrated with a workshop script that guided the users through the tasks they had to accomplish (see Appendix A). After every task, a series of questions was asked to gather feedback from the users. Questionnaires are a good way to measure a user’s subjective satisfaction and preferences, which were noted earlier as two central elements of a usability evaluation. An alternative method to get feedback from the users is to do individual interviews. However, interviews require more time and staff than were available for this thesis. Given the limited resources in the project, questionnaires were chosen to survey a larger number of people. A drawback with this approach is that ideally a questionnaire should undergo a pilot study to refine the questions and prevent misunderstandings in the data collection. Furthermore, while a questionnaire can provide rich subjective feedback, it is also subject to different types of bias and, as Nielsen (1993, p. 209) points out, “One cannot take user statements at face value. Data about people’s actual behaviour should have precedence over people’s claims of what they think they do”. Thus, questionnaire data need to be complemented by a method that collects objective data on the users’ actual work. To achieve this, classic usability engineering methods were used by logging user interactions with the software. In other words, the software automatically recorded all actions that the users performed with the interface. Each action occurrence was time-stamped so that a sequence of actions could be recreated and statistic counts on user actions could be calculated. Software logging is particularly interesting as it can run continuously in the background without interfering with the user, and it allows for easy identification of frequently used or unused features. On the other hand, it can violate privacy of users and therefore they have to be warned that their interactions are being logged. This was done by an introductory letter and a form of consent that were provided with the workshop materials. Table 4.3 summarizes the combination of complementary data collection methods used for this study.

Data collection methods	Type	Pros	Cons	Application
Direct observation	Qualitative Subjective	- Easy to perform - Suggests functions and features	- Obtrusive - Incomplete	Selective
Thinking aloud	Qualitative Subjective	- Information on cognitive activity - Find user misconception	- Add strain on the participants	Selective
Software logging	Quantitative Objective	- Unobtrusive - Run continuously	- Difficulties to analyse important amount of data - Ethical issues	Systematic
Questionnaires	Quantitative/ Qualitative Subjective	- Easy to repeat - Find user preferences	- Pilot work required - Bias answers	Systematic

Source: adapted from Nielsen (1993) and Preece et al. (1994)

Table 4.3: Summary of the data collection methods

4.5 Workshops

4.5.1 Sampling Strategy and Recruitment

To perform usability tests, a sample of users had to be selected and recruited. These individuals had to be as representative as possible of the intended users of the software (Nielsen, 1993). As noted earlier, the potential range of users for MapChat Viz is very wide and can vary from people who are unfamiliar with computers to GIS experts. Therefore, the test user group ideally had to comprise people with different computer skills from novice users, to ensure that the features are easy to understand and used, to expert users, to provide insights and qualitative feedback on the system. To meet this requirement, two subpopulations of users were identified, namely the graduate and undergraduate students from the Faculty of Environment at the University of Waterloo and the participants in the previous MapChat workshops in the Bulkley Valley described in Section 4.3.1. The student group was expected to be familiar with computers, the Internet, and GIS. They were mostly aged 20-30 with a roughly even gender split. The second group of test users were expected to be less technically knowledgeable than the first group and represent a higher proportion of non-expert users.

The participants from the Bulkley Valley were also important in this instance as they already had an experience with MapChat software, most of them were involved in the planning process of the area and most were therefore familiar with some of the assets recorded in the database. This group was mostly aged between 40 and 60 with a roughly even gender split. These two groups were intended to cover a wide range of the intended users of the system.

Non-random sampling techniques were used to recruit the candidates. Purposive sampling coupled with snowball sampling was used to recruit participants amongst these two groups (Sheskin, 1985). The researchers from UW were in contact with the leaders of the BVSC who maintained contact with the participants of the MapChat community asset workshops held earlier. Thus, an invitation to participate to the MapChat Viz workshop was sent out by email to all previous participants. This was followed by a reminder email. Students were recruited from the University of Waterloo via an invitation sent out to the Faculty of Environment graduate student mailing list. Also, two presentations were made in undergraduate GIS courses. In both cases, potential participants were mainly contacted by email. This was found to have some limitations, for instance some of the people who participated in the asset workshops in the Bulkley Valley did not have an Internet connection and thus were not contacted. Hence, the recruitment has a bias towards people who were already using the Internet and were at least casual users of the Internet. Table 4.4 summarizes the recruitment process and numbers, not surprisingly, the response rate for the residents from the Bulkley Valley who were previously involved in research and planning initiatives in the study area was much higher than for University of Waterloo students, for which the main motivation factor related to curiosity of a fellow student's research project.

Place	**University of Waterloo**	**Bulkley Valley**
Number of people emailed	~ 350	40
Number of answers	14	17
Number of participants	9	13

Table 4.4: Summary of the number of people through the recruitment process.

4.5.2 Workshop Procedure

As implied earlier, the MapChat Viz case study took place through a series of workshops located in two places, Waterloo, Ontario and Smithers in the Bulkley Valley, BC. In both locations, participants were provided with several dates for group workshops and, if they were not able to come at those dates, individual workshops were arranged. The same materials were presented for the individual and group workshops, however there were a few differences between the workshops located in Waterloo compared to those in Smithers, as is explained later. Overall, there were two group workshops held in Waterloo with two and seven participants, respectively one group workshop in Smithers with 12 participants, and one individual workshop in Smithers for a total of 22 participants. The complete study and its associated materials were reviewed and validated by the University of Waterloo Office of Research Ethics.

The workshops took place in two different computer labs including the spatial decision support lab in the Faculty of Environment at the University of Waterloo, and the computer lab of the Northwest Community College in Smithers. Both labs had Internet access and computers able to run a Web browser efficiently. The workshops were conducted by one facilitator in Waterloo and two facilitators in Smithers. Upon the arrival of participants at the computer labs, they were each given an introductory letter that described the workshop procedure, a consent form required for ethics approval, and the workshop script. They were invited to choose a computer to use in the workshop. Once the introductory letter was read and the consent form signed, the workshops in Waterloo were opened with a quick group presentation of the study area, the data collected in the previous community asset workshops, and the issues at stake in the case study. In the Bulkley Valley, the workshop was more flexible to accommodate the participants' personal needs. Thus, participants could arrive at the computer lab during a specific window of time. This format did not lend itself well for a formal group presentation, however most participants were already familiar with the broader PPGIS research initiatives in the Bulkley Valley. Thus, the participants were given a brief individual presentation of the specific goals of the workshop to complement the introductory letter. Participants were then guided by the workshop script to connect to the Web-based computer program MapChat Viz hosted at the University of Waterloo.

Following these introductory procedures, the participants used the workshop script to guide them through the series of tasks described in Section 4.5.3. After each task, the users were asked to answer a short set of questions included in the workshop script. To prevent the users from being overwhelmed with the new tool, the interface was presented incrementally by adding the features when needed for a new task. According to the methods presented in Section 4.4.4, facilitators walked around the lab observing the users and taking notes without interfering with the participants. If the facilitators were asked for help, they asked the users to formulate their action and their issues as recommended in the “think aloud” protocol. Comprehensive help documents, online and hard copy, were also provided for the users to refer to. However, only a few users actually referred to the help documents as most preferred to ask the facilitators for assistance. As mentioned in Section 4.4.4 , the interactions between the users and the software were continuously logged. At the end of the session, an open discussion took place to answer any questions from the participants. Finally, an appreciation letter was sent to all the participants after the workshop was completed.

4.5.3 Workshop Script and Questionnaire

The workshop material in Appendix A integrates a script to guide the participants through the steps and tasks they were asked to perform with a questionnaire to gather feedback after each task was completed. The workshop material is organized into 5 sections (see Figure 4.6). Each section contains a task or a set of tasks that were designed to test a feature or group of features related to a particular issue aspect of the software. The length of the workshop and the time required to complete it is an important practical constraint on the workshop format. Indeed, only a reasonable amount of time, between an hour and an hour and a half, can be asked from volunteer participants (Nielsen, 1993). Thus, the questionnaire and the associated tasks had to be constrained to be completed within these time limits. Furthermore, the tasks had to be as realistic as possible in order to increase the users’ understanding and also to make the workshop tasks close to a realistic application of the software. One way to achieve these goals is to relate the tasks to an overall scenario (Nielsen, 1993). In this case, the tasks were based on a scenario where participants assumed to be community leaders who were going to attend a town meeting and wanted to build a case to defend some amenities of particular importance to the community.

Questions had to be as unambiguous as possible to avoid misinterpretation that would bias the results. A combination of closed and open questions was used in the design of the questionnaire in order to find a balance between quantitative and qualitative feedback from the users. Closed questions can include multi-point rating scales, Likert scales, and semantic differential techniques (Preece et al., 1994), which allow for numeric codification and statistical analysis. Open-ended questions were used to gather more in-depth explanations from the users. While open-ended questions are an economical method to collect qualitative feedback, they require more involvement from the test users who have to transcribe their thoughts onto the paper. Because of the time constraints of the project, a comprehensive pilot study was not conducted to calibrate the duration of the workshop and avoid ambiguous questions. Only a few test runs were performed.

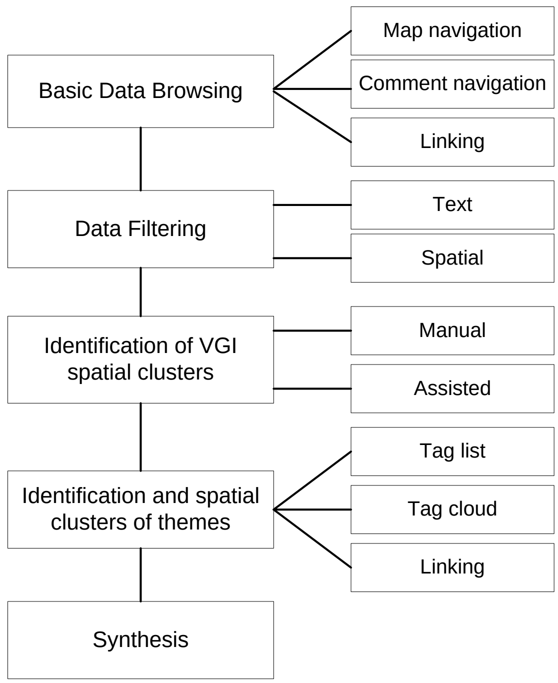


Figure 4.6: Overview of the workshop structure

The first few pages of the materials in Appendix A consist of the invitation letter and the introduction describing the workshop aims and procedures. Before operating the software, participants answered a series of questions about their background including some demographic information; experience in using computers, the Internet, and GIS use; familiarity with the Bulkley Valley area; and their involvement in local planning. Data about user skills with computers and GIS allows verification that the sample constituted a wide range of users and also to identify differences in opinions between experts and novice users. Knowledge of the Bulkley Valley and the participants' involvement in local planning allowed us to study the influence of these parameters on the use of the software to be evaluated.

In accordance with good usability practice, the first section of the process involved relatively easy tasks to increase user confidence and allow the participants to get familiar with the new application and its basic features (Nielsen, 1993). The first part of this section invited the user to try out, through small descriptive steps, the basic tools available to explore the map (map tool bar: zoom in, out, panning...) and the user comments (popup, comment grid) as well as some basic linking between map drawing and text comments and brushing methods. After getting familiar with these features and providing feedback, participants were asked to achieve more precise goals through Task 1 and 2. Task 1 was designed to make the users look for VGI features within a geographic area (e.g. “find three features near a river”) whereas Task 2 made users look for VGI features related to a specific word (e.g. “find three comments related to trail”). At this stage, the participants had access only to the basic features of MapChat Viz. Therefore, the overall expected outcome of this first section was that most users would find it easy to navigate the map with the basic tools but their navigation would be hampered by the volume of data in the comment grid and on the map when they tried to find specific information.

In the second section of the workshop instructions (see Appendix A), the text filter and spatial filter were added to the interface and the users were asked to repeat slightly changed versions of Task 1 and 2 with the help of these new features. Two closed questions were used to determine if the user actually found the filters helpful when achieving the tasks. Two open questions were asked to evaluate the ease of use and understanding of the two features. It was expected that the users would relatively easily understand the features and take advantage of them to aid exploration of the VGI contained in the map viewer and comment list.

The third section of the instructions (see Appendix A) introduced the clustering tools designed to help the users browse the data and to visualize areas with varying density and low density of VGI (see Section 3.3.5.4). To evaluate this tool, the participants were asked to locate three areas that they thought have a high density of drawings and circle them with the drawing tools. They achieved this task once manually and once assisted by the clustering tool. The shapes they drew for their answers were saved in the database for both tasks in order to compare the areas chosen and to determine the influence of the tool. A set of questions were asked to determine if the users thought that the clustering function improved the legibility of the map and helped them to locate areas of concentration. Another set of questions were related to the users’ confidence in the system and its understanding and the complexity of the feature. These points are important as the system must be transparent and not a black box in order to support an informed review and evaluation of the collected VGI. The feedback expected on this feature was that the tool was easy to use and that it provides valuable information on the spatial partitioning of the data. However, it was also expected that the aggregation may be seen confusing and that visual effects could be improved in order to enhance its understanding, especially to show what features are included in what cluster.

The fourth section of the instructions addresses the issue of non-categorized data. To examine this, each comment was tagged with one or several theme tags and a tag list and a tag cloud allows users to visualize the data categories, their relative frequency, and their spatial location (see Section 3.3.5.5). The main objectives here were to assess the ease of use and usefulness of the overall tool as well as to compare the specific tag visualization features. One task was designed to provide the users with a realistic context to experiment and compare the various features. Participants were asked to determine an area of interest with the drawing tool. The choice of the area was left to their discretion. They had then to determine and report the themes that were visible and their relative frequency within the area of their choice. Participants were then asked to evaluate the usefulness and complexity of each tag visualization features as well as to compare them. The classification and the related visualization tools were expected to be relatively easy to use and valuable to explore the data content. In this case, the tag list was expected to be the most appropriate visualization feature as it seems to be simpler to interpret.

The last section of the workshop comprises one large task that was designed to put the participants in a less structured setting and within a more realistic context that sought to capitalize on all of the features learned earlier in their use of the software. It also aimed to make participants feel rewarded for their participation and accomplishments (Nielsen, 1993). To achieve this, the participants were asked to pick a theme of their choice and then to identify areas related to this theme that they would prioritize for special attention in the community. The identification of these areas was done by defining areas on the map. The resulting shapes were recorded in the database for subsequent analysis.

The outcome of such a task is interesting on several levels. First, as mentioned by Andrienko et al. (2002), understanding how to use a tool, its purpose and when it is useful are three different levels of understanding. Thus, confronting the participants with a broader task allows determination of whether the participants understood the tools and their relevance beyond the scripted use. Second, it is useful to push the limits of the tools in order to discover unexpected uses that can be used to refine task definitions for future studies as well as the limitations of existing features. Finally, it also provides a basis to compare the usefulness of the different features in the data exploration process thanks to the feedback provided by user ratings in questionnaires and the software interaction logging. The hope was that the users would take advantage of the full set of tools, but it was expected that most users would only use a subset of the simplest tools as they did not have enough time to grasp the complete relevance of each tool. The hypotheses related to the result of each task are summarized in Table 4.5.

VGI visualization issues	**Technique implemented**	**Workshop script tasks/questions**	**Hypothesis**
Simple Browsing and linking	Highlight, Popup, zoom to and brushing	Section1: Task 1 and 2 Question 1,2,3,4	- Popup is well understood to ensure access to the information linked to a given geometry - “Zoom to” is sufficient to ensure the location of a geometry linked to a given comment - Brushing ensures a quick browsing of the data - Tools are satisfactory for random browsing but are limited to find specific features
Browsing and Linking	Spatial filter	Section 2: Task 1 Question 1 and 3	- Spatial filter improve the access to the comments related to a given area
Browsing and Linking	Text filter	Section 2: Task 2 Question 3 and 4	- Text filter improve the location comment related to a word topic
Browsing and Heterogeneous data	Clustering	Section 3: Task 1a and 1b Question 1 to 9	- Clustering improves the readability of the map - Clustering helps users to identify areas of concentration - Clustering symbolisation is relatively well understood - Clustering process may not be well understood
Non categorized data	Tagging, tag cloud, tag list, Clustering pie chart	Section 4: Task 1 Question 1 to 6	- Tagging and related visualization features improves the theme browsing - Tag list is easier to use than tag cloud
Comparisons of the methods	All the features	Section 5: Task 1 Question 1 to 3	- Users focus on the use of one tool rather than a combination of tools

Table 4.5: Summary table of the issues, features, hypothesis and outcome

4.6 Chapter Summary

This chapter described the study area and the process that was used to collect a dataset that is representative of Type 2 VGI. The discussion turned next to how this dataset was loaded in the prototype application in order to allow participants to use and evaluate the visualization tools with actual VGI data. In order to collect data on the use and evaluation of the tools by the participants, a set of research methods from the usability engineering field were examined. To mitigate some of the limitations related to the use of standard usability methods in the context of geovisualization, an approach combining several of these methods was employed to collect complementary forms of usability data. This approach was put into practice in a series of workshops where groups of participants used and evaluated the visualization tools.

The next chapter presents the analysis of the data collected during the evaluation workshops with the aim to assess the influence of the geovisualization techniques on the participants' abilities to explore and extract meaning from the community asset data.

Chapter 5
Results

The series of workshops organised for the case study led to the collection of both qualitative and quantitative data concerning the participants' use of and satisfaction with the MapChat Viz tools. These data provide a solid base to examine the relative merits of the tools and methods in the software with respect to the enhancement of VGI exploration and visualization. The first section reviews the characteristics of the participants. The chapter is then divided in five sections that discuss the results and findings related to each set of visualization features based on the analysis of the data collected.

5.1 Participants

Understanding the participants is especially important when evaluating any software since usability issues can vary with users of different characteristics (Nielsen, 1993). Moreover, although participant recruitment typically aims to reach test users who are representative of the target audience of the software, lack of resources and other study constraints can affect the composition of the test user sample. While Section 4.5.1 described the goals and methods of the participant recruitment, this section review the characteristics of the actual participants to determine if the recruitment process was efficient and if there are any biases that could influence the results. In this process, it is also important to delineate subgroups of participants with similar characteristics to determine how the different types of participants respond to the software. A total of twenty-two participants were recruited, nine participants for the University of Waterloo workshop and thirteen for the Bulkley Valley workshop. The participants are characterized below based on their answer to the questionnaire preliminary section (see Appendix A, Introduction).

A participant's age has been recognized as a general indicator of their familiarity with and their ability to understand and to interact with geospatial technologies (Slocum et al., 2001). This study focused on participants aged 18 years and older to represent the heterogeneous population that uses VGI applications. Deliberate efforts were not made to recruit children, young adult and seniors, however it would be interesting to include a wider age range of participants in future studies on VGI collections, especially because their views and ideas can be significantly different from the main adult population. For this case study, the recruitment resulted in a participant sample with an even age split across the target age range (see Figure 5.1). This age distribution is suitable to represent the targeted panel of users.

Although the influence of gender on the interaction with maps has been extensively discussed in the literature, findings are very diverse and have overall mainly shown that only the gender variable "cannot predict any individual's competence at a geospatial technology task" (Davies, Chao, and Albrecht, 2010, p. 30) (Meng and Malczewski, 2009). Therefore, even though the influence of the gender variable is complex to determine, the characteristic was recorded to ensure that the participants are representative of the target user population. The gender split of the recruited participants is slightly skewed towards the male gender that represents 59% of the participants because of the high proportion of male students in GIS studies that overcomes the even gender split of the Bulkley Valley participants.

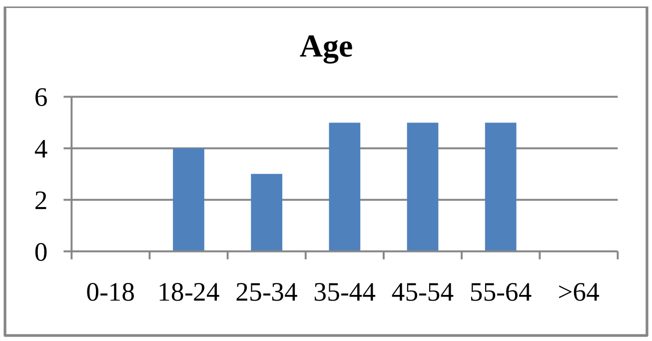


Figure 5.1: Gender and age of the participants

As described in Chapter 4, the intent was to recruit participants with various skill levels and, hopefully experiential knowledge. Engaging people with various educational and computer backgrounds is a recurrent issue of studies related to public participation and special means have to be employed to reach populations that are otherwise excluded (Meng and Malczewski, 2009). Due to the limited resources available, the recruitment of participants was done mainly by email and relied on the personal interest of the people for the Bulkley Valley and/or GIS technologies to prompt them to participate to the workshop. As a result, the participants' familiarity with computers and maps was not as heterogeneous as expected in the wider population. Two people can be considered as not at ease with computers and the Internet whereas twenty people are regular computer and Internet users (see Figure 5.2). Concerning computer-based map use, the distribution is a bit more spread out with five people who can be considered as not familiar with mapping software versus seventeen who operate mapping software frequently. Even though the number of non-expert users is relatively low, the overall sample of participants represents a wide of range of users that can be divided in two subgroups: a non-expert group of five people and an expert group of seventeen people.

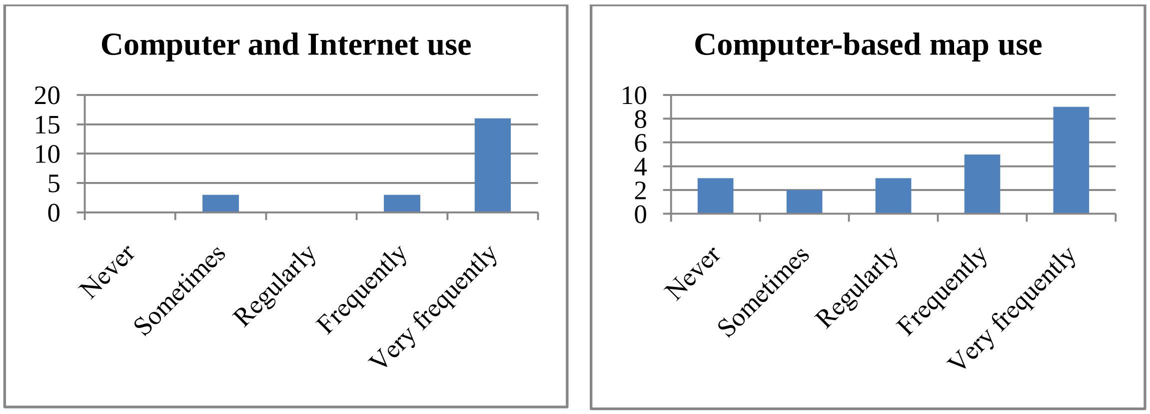


Figure 5.2: Computer, Internet and mapping use of the participants

Another important characteristic of the participants that can influence the feedback they provide is their familiarity with in the Bulkley Valley area and their involvement in its planning processes. As described in Chapter 4, participants were recruited from the University of Waterloo Environmental Studies student body and the group of Bulkley Valley residents who participated in previous workshops. As expected, given the small odds for people from Waterloo to know a secluded area located thousands kilometres away, the number of people familiar with the Bulkley Valley corresponds to the number of people who attended the workshop in the Bulkley Valley: thirteen are familiar with the Bulkley Valley and nine are not. Within the people from the Bulkley Valley, it is interesting to note that everyone was involved in the local planning of the area and only one person was not at the previous MapChat workshop. Practically, the distinction between participants familiar with the Bulkley Valley area and the others was difficult to study. The expert/novice characteristic was an important confounding factor as all novices were in the Bulkley Valley group (see in Table 5.1) and so the main differences that were found between the two groups were most likely to be attributed to the expert/novice distinction since the computer experience is likely to have more impact on the experience with the software than the familiarity with the area. Therefore, the distinction between participants familiar with the area and non-familiar was eventually left aside in the presentation of the results.

User	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Origin	UW									Bulkley Valley												
Level	Expert									N	Ex	N	Ex	N	N	Ex	N	Ex	Ex	Ex	Ex	Ex

Table 5.1: Repartition of expert and novice users between the two locations

The workshop consisted of a 5-section script with each section was closely linked to one of the tools implemented in MapChat Viz. Each section included testing tasks that were followed by several questions and opportunities for feedback. All the participants entirely completed their questionnaires with the exception of a few questions missed accidently in a few instances and one questionnaire that misses the last part due to the time constraints of the participant. The completion results were not as good concerning the shapes that participants were asked to draw to complete some of the tasks. Only twelve participants out of twenty two saved their drawings correctly for task 3a and 3b, thirteen for task 4 and nineteen for task 5 (see Table 5.2). The main assumption to explain the errors in task 3 is a mismatch between the succession of the tasks written in the script and within the software. Despite the fact that the participants were carefully guided in the script on when they had to go to the next part or save drawings, and that the numbering of the task corresponded between the screen and the script, it appears that it was not enough and drawings relative to one part were saved mistakenly for another part. Some participants may have also just forgotten to save their drawings despite the bold reminders in the workshop script. In the following sections of this chapter, the number of answers taken into account to support each analysis will be mentioned to clarify their reliability.

5.2 Simple Browsing

5.2.1 Tool Bar

In order to visualize and explore VGI, the user must first be able to interact with VGI in basic ways. For instance, the user has to be able to change the scale or the extent of the visualized data as with any geographical data. Traditionally in desktop GIS, the primary way to enable the users to browse geographic data has been one or several tool bars with series of icons. While popular Web mapping sites such as Google Maps or Yahoo! Maps offer simplified interfaces, many Web mapping applications have an importance inheritance from desktop GIS and have simply transposed the tool bar from the complex desktop GIS to the Web environment taking for granted its usability (Noyle and Bouwman, 2009). However, it appears that this approach requires more investigation as the audience of Web mapping applications often includes non GIS professionals for whom it is not as easy to use classic tool bars as reported by Kramers (2008, p. 102) in his usability study of the Atlas of Canada Web application:

> "Their functionality was based on similar GIS tools and the team expected that users would understand the behaviour of the tools. The results of the first survey indicated that users came with different expectations and understanding of how tools would function. In many cases, they had little previous online mapping experience to guide them. Their understanding of mapping and interacting with online maps was considerably less than what was assumed by the design team. The result was that many users could not effectively use the tools to explore the Atlas' maps."

The investigation of the tool bar was not the primary goal in the MapChat Viz study so the default tool bar provided by the development tool kit was implemented (see Figure 3.6). Even though no task was dedicated to the tool bar and no quantitative feedback was collected about the basic navigation tool bar, some participants especially the novice users who are not used to GIS reported that the tool bar icons were too small to easily click on or to read. As a consequence, it was difficult to identify their function from their icons and to determine when they are highlighted or not (i.e. activated or not) similarly to Kramers (2008) findings. These difficulties were worsened by the default behaviour of the tool bar that toggles tools on and off. When they clicked several times on a tool icon to make sure that it was activated, the tool got deactivated and the default tool (panning tool) was activated which caused confusions for the users.

These findings confirm that taking for granted the fact that GIS desktop conventions are known by all users is generally a misconception. Basic features like the tool bar are often overseen and should be adapted for various users who are not familiar with GIS. (Harrower and Sheesley, 2005; Kramers, 2008; Nivala et al., 2008). Several simple improvements such as increasing the size of the tool bar icons, keeping a tool activated when it is clicked on, changing the mouse cursor to display the icon of the tool that is active (i.e. a magnifying glass for the zoom in tool), adding plain text besides the icons as well as verbose tool tip (Kramers, 2008) and making tools that accommodate different types of users cohabite in one interface (Harrower and Sheesley, 2005; Kramers, 2008).

5.2.2 Spatial Browsing and Information Windows

In the first section of the workshop script (see Appendix A, Section 1), participants were guided through a few steps to get familiar with some basic spatial browsing tools such as the map navigation and the information windows. Browsing spatial features on a map and enquiring for information about a specific object is a basic task of geographical data exploration in general and especially VGI. The info window or pop-up window is an easy solution to allow users to display information on request about a specific object and to link the geographic data to its related alpha-numeric content (Sheesley, 2009). Given its wide implementation in Web mapping and other fields, it was expected that info windows would ensure a quick access to the information related to the drawings in a usable and effective manner for a wide range of users. However, it was questioned if the default info window, that has almost become an imposed standard through the popularity of Google Map, was optimal given the many possible variations on the size, position, look, behaviour that have rarely been experimented (Sheesley, 2009). It was also expected that info window would show some limits in usability and efficiency when users have to achieve tasks involving the browsing of numerous features in specific areas of interests, a common task of VGI exploration.

After following the simple steps to get familiar with the software, the participants were asked to provide quantitative ratings on the ease of use of the pop-up tool to display comments related to a drawing(s). As expected, the quantitative feedback about the info window tool was very good (15 users found it very easy to display a comment with an info windows, 6 found it easy, and 1 neutral) but several improvements were suggested through qualitative feedback. Three remarks were directly about the pop-up tool including two on the opening/closing management. In the current state, pop-ups can be closed manually by clicking on the "X" sign in their top right hand corner or automatically by opening a pop-up on another drawing. While some users suggested that any other click away from the pop-up should also close it, it appears that the behaviour of the pop-up is not that simple (e.g. a user may want to drag the map to get more geographic context while keeping the pop-up open) and should be carefully studied for each of the potential actions. The behaviour implemented in the application appears to be a good compromise (Sheesley, 2009). Some participants also suggested making the pop-up draggable to allow the users to move it at their convenience enabling them to browse the drawings that are obstructed by the pop-up. This suggestion is consistent with the improvement suggested by Sheesley (2009) where draggable info windows can allow several pop-ups to be open simultaneously. This permits data related to several drawings to be compared without obstructing the relevant part of the display. However, it is easy to lose track of the relationship between the multiple windows and the drawings so ways to maintain strong visual linkages between the info windows and the locations without overloading the display need to be investigated (Sheesley, 2009).

Four other remarks were related to the dynamic linking of the drawings towards the comments displayed in the comment tab (see Figure 3.7). As presented in Section 3.3.5.2, the comment tab is dynamically linked to the drawing through the brushing and “zoom to” features. However, when clicking or hovering a drawing, the related comment is not highlighted in the comment tab. At the time of design, it was deemed that the pop-up tool would be enough to ensure this function and that it would be confusing for the users if the comments kept scrolling up and down while browsing the drawings. However, it appears that at least four users indicated that it would be better if the comments corresponding to a drawing would automatically show in the comment tab. Therefore, it seems that brushing and linking from the map view toward the comment tab view should be implemented to establish as a two way link accordingly to the coordinated multiple view practices.

In the second part of the first section of the workshop, users were asked to browse three comments related to drawings nearby a river. The aim was to explore the limitations of the pop-up tool to spatially browse numerous features in an area of interest. However, the results were not as significant as expected. Seven participants reported no difficulties at all to achieve the spatial browsing task and fifteen made suggestions to improve the overall spatial browsing experience without pinpointing the expected limitations. In retrospect, these results are mainly due to the simple nature of the task assigned to the users since finding only 3 features was not likely enough to expose the limitations of the approach. This task was designed to limit the time asked of the volunteers but a higher number of spatial features should have been used.

Finally, three participants reported that it was difficult to find or visualize features either because of the lack of landmarks or because of overlapping features, especially when users had no prior knowledge of the area. These remarks confirm the importance of landmarks as a mechanism for the users to contextualize the map (Jones, 2010). The background map layer only had labels for some of the street names and landmarks so adding more information on the map could be a solution. However, this has to be done carefully because the display can be easily overloaded and moreover, getting data for a secluded area is difficult. In this case, some Type 1 VGI could be useful to serve as base data but no datasets were developed enough in this area at the time of the case study. Also, some labels on the background layer were obstructed by the VGI layer so it would have been helpful if users had an easy way to turn this layer off without adding complexity to the application. The impact of overlapping features was already limited by the implementation of highlighting based on the colour and the depth of field on mouse over (see Section 3.3.5.1), but it could be improved by combining the existing highlighting method with one based on transparency where not only the object of interest is highlighted but others around are faded (Robinson, 2006).

Overall, the info window has proven to be an easy and intuitive tool for simple information probing but further experimentation on the more advanced pop-up design and behaviour is needed to improve its efficiency of use. Such experimentation could be facilitated by features recently added to the GeoExt tool kit that offer draggable pop-ups out of the box. Even though the results did not reveal the info window limitations as much as expected due to the limited scope of the task, the info window appeared to be limited to explore numerous VGI features.

5.2.3 Comment Browsing and Location of Related Drawings

The first section of the workshop also included steps for the participants to get familiar with the tools involved in finding relevant information and determining where the related drawing is on the map, the counterpart of spatial browsing. Users can explore comments by scrolling the comment list up and down. When they identify a comment of interest, they can click on it to have the display centre and zoom on the related spatial feature or they can simply move the mouse over the comment to highlight the drawing on the map. The “zoom to” feature is commonly found in GIS and linking and brushing have become standard features of geovisualization and data exploration tools overall. However, they have yet to be tested more widely in Web mapping applications used by various types of users and it is thus important to determine their usability and effectiveness in this context. It was expected that these two features would be intuitive for the users and useful for random exploration but limited for targeted exploration.

After familiarizing themselves with the software, participants were questioned on the ease of use of the comment tab to locate a drawing(s) related to a given comment. This task was rated as easy by 6 participants and as very easy by 16 participants. This was confirmed by the observations and users’ remarks on the “zoom to” tool that was easily taken in hand by all the users. One small inconvenience that was observed is that the software uses a fixed zoom level (or scale) for all the drawings. This can give variable results depending on the size of the drawing and the intended meaning of the drawing. One simple improvement would be to calculate the zoom level based on the feature size, however this might fail to convey the meaning that the contributors originally had when they drew the features. Indeed, the accuracy or vagueness of the drawing is linked to the scale at which it was drawn and this scale may not match with the one that would be automatically calculated (De Longueville et al., 2009). Therefore, a better solution would be to record the scale at which each of the drawings were originally drawn and use it as a reference for the “zoom to” feature for each feature. This way, the feature “zoom to” would be easier to use and the potential to misinterpret other people’s drawings would be reduced somewhat.

The user ratings and facilitator observations were contradictory concerning the brushing and linking features. It appears that the brushing feature was not obvious for all the users and that many relied almost exclusively on the “zoom to” feature. Indeed, the facilitators observed several times some users who were surprised when the brushing feature was explained to them at a later stage of the workshop. To make the brushing more obvious for some users, a different highlighting technique could be used such as the leader lines technique that visually link the comment and the related drawing together on the screen by using lines (Robinson, 2006, 2009). Users who find leader line visualization too intrusive should be able to switch between different highlighting techniques.

Several users who used the brushing feature suggested one major improvement through their qualitative feedback. The brushing was only implemented on mouse over also referred to as transient brushing since the highlighting of the linked drawing lasts only while the cursor is on the comment. This was found to be confusing by several users as they would highlight a feature and then try to move the cursor to it on the map and the highlighting would disappear since the cursor was no longer over the comment. To address this issue, the transient brushing can be completed by a durable brushing that consists in triggering the highlighting when clicking on the comment and keeping it on until another comment or the same comment is clicked on as implemented by N. Andrienko et al. (2002).

Finally, in the second part of the first section of the workshop users were asked to find a specific type of features (in this case trails) by using the simple list of comments. Thirteen participants reported that the task was tedious and suggested that a search function that highlights keywords in the comments would be more useful. As these results were expected, the next section was designed to determine the enhancements provided by the text filter and the filtering functions in general.

5.3 Interactive Filtering

5.3.1 General Findings

In the second section of the workshop, interactive spatial and textual filtering functions were introduced to the users and added to the interface (see Section 3.3.5.3). The users were then asked to repeat two tasks similar to the ones achieved in the first section (i.e. finding a specific type of comments and finding feature in a given area) but this time with the help of the new tools. The objective of the section was to evaluate if these tools mitigate some of the limitations encountered by the users in the first section where only basic features were available. A parallel goal was to assess the usability of these features for the various users. It was expected that the filters would improve significantly the data exploration experience for the users without major usability issues. It was also anticipated that the text filter would be easy for everyone to understand and use since such functions are similar to what is found in Web search engines. Before presenting the specific results for each type of filter, some general findings about the filters' implementation are presented.

A major finding that emerged from various observations and participants` feedback is the need for the software to provide more visual cues concerning the current status of active visualization features such as filters. Such information can be about the content of a view, the status of a view (for instance if a view is being refreshed or has been refreshed), the parameters used to generate a view (for instance if a filter is being applied to the view content), the linkages between the different views, or any other information that is difficult for the users to remember while interacting with the tools. These pieces of information help the users to use the tool more efficiently as well as they avoid potential misinterpretation of the data visualization. The development of techniques to display contextual information about the visualization constitutes a nascent theme of the geovisualization research referred to as meta-visualization (Roberts, 2008; Weaver, 2004, 2005).

In the case of MapChat Viz, some users were seen to misinterpret the data they visualized because they had forgotten that the spatial filter or text filters were enabled. Although it was possible for them to check if the spatial filter box was ticked or if a word search was entered, more obvious visual cues are required. To make this information more apparent and understandable, it would be interesting to use a watch window that would display and centralise in one place a summary of the information on the status of the filters along with other relevant information such as the geographical extent that is being used for the spatial filter or the word(s) currently used by the text filter (MacEachren, Brewer, and Steiner, 2001).

Moreover, several users reported that they would like an explicit indicator that notifies when the comment list is being refreshed. Currently, when a filter parameter is changed (e.g. new spatial extent or a new word search), the user cannot determine if the content has been refreshed or if the content is obsolete and is being refreshed. This can be fixed by adding a symbol such as a spinning wheel over the comment list that notifies the user that a calculation is being performed. Providing informative feedback from the system to the users is one of the usability golden rules so this issue should have been avoided by doing a thorough usability heuristics study that was not done due to the time constraints of the project (Haklay and Nivala, 2010; Nielsen, 1993, 2005).

The need for such meta-visualization techniques becomes even greater for advanced use. For instance, an expert participant reported that using the two filters in combination was very useful. However, it has been observed that most novice users got confused when having both filters activated. In this case, the understanding of the combination of the features could be significantly eased through meta-visualization.

5.3.2 Text Filter

To evaluate the merits of the text filter tool for comment browsing, the participants were asked to find three comments specifically related to wildlife. Text search tools have become widespread especially through the use of search engines. Therefore, as expected, the comment browsing task with the help of the text filter was found easier by 4 participants and much easier by 18 participants. However, since many participants are used to advanced search tools that permit searches based on several keywords and optionally logical operators, they reported that MapChat Viz text filter was too limited. The software logs actually testify of the user attempts to perform search such as “wildlife deer” or “wildlife + deer” that did not return relevant results in the current state of the text filter tool. Although, the text search was implemented in a basic form due to time constraints on software development, these findings show that a classic search engine approach should be implemented. Other smaller improvements requested by participants include highlighting the keyword within the comments and providing an indicator of the total number of comments retrieved by the search.

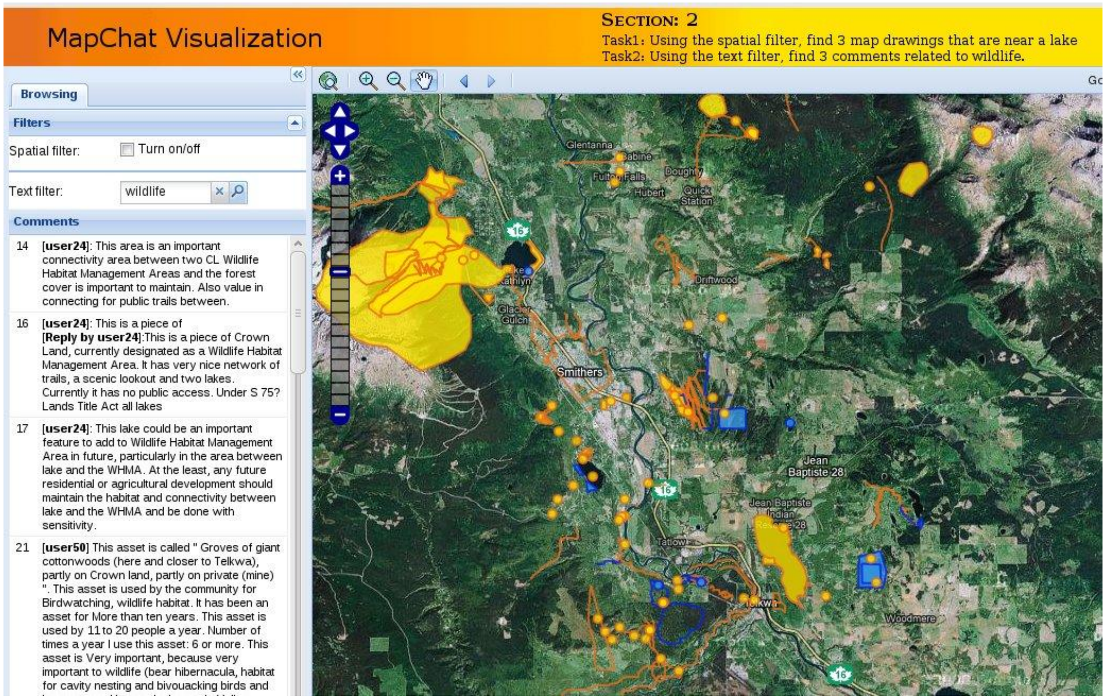


Figure 5.3: Drawings highlighted for a search on the keyword “wildlife”

Beyond the simple comment browsing, the linkage of the text filter and the map through the highlighting of the drawings related to the comments retrieved by the search was found very useful even by novice users due to its visual intuitiveness (see Figure 5.3). One novice user noted that “it is excellent for impact analysis”. It therefore confirms the importance and usefulness of using coordinated multiple view techniques such as the brushing and linking to improve the exploration of the spatial and thematic relationships between the multiple contributed comments and drawings.

5.3.3 Spatial Filter

To experiment with the spatial filter tool, the participants were asked to identify several drawings located in a given area, in this case near a lake. Spatial filter functions are not as common as text filters and moreover it involves spatial concepts specific to mapping applications. It was thus expected that the spatial filter would be more difficult for some people to understand compared to the text filter. Surprisingly, the spatial filter received a good rating from all users. The spatial browsing task with the help of the spatial filter was found much easier by 15 participants, easier by 5, and the same by 2. However, these results are somewhat suspect since the facilitators observed that several participants did not take advantage the spatial filter to complete the spatial browsing task more efficiently but instead they opened the info window on each of the drawings they reported as they did in the first section and did not see the utility of the spatial filter. These observations are confirmed by the software logs. All of the users did activate the spatial filter and experimented with it as the workshop script requested. However, thirteen users opened the pop-up on each of the features they reported in their answer to the task whereas only five participants including four expert users did not use the pop-up at all and four used the pop-up once out the three features they were asked to report. This divergence between the rating and the actual use of the software can be explained by a bias of the respondents who trying to answer based on what they think the facilitator would want rather than a judgement purely based on experimentation. While the quantitative rating results need to be mitigated, seven positive qualitative feedbacks clearly showed that they understood how to use the tool and specifically mentioned the easiness and intuitiveness of the tool.

The fact that some users did not use the spatial filter is partly due to the problem in the task definition identified earlier. Since the task was too limited in scope, most participants did not feel compelled to experiment with the spatial filter to complete the task and preferred to use the basic exploration feature such as the info window and comment list. Beyond the issue of the ill-defined task, this shows that the feature could be made more useful and intuitive for novice users. As mentioned in the general finding section, more visual information is needed to inform the user of the filter status and context. For instance, a watch window could display a flashing red ribbon that informs the users when the spatial filter is active.

Another explanation for the limited use of the spatial filter is the way the spatial extent is defined. During the software design, the decision was made to use the extent of the screen display as the extent of the spatial filter to make it easier to use. This ease of use was confirmed by seven qualitative answers. However, it may not be the best solution in all cases. Indeed, a user reported that the spatial filter was only useful when zoomed in but not really useful when zoomed out because of the numerous comments in the list that makes it inefficient. Another more minor issue reported by a participant is the fact that some actions such as “zoom to” or opening a pop-up move the map display which makes the users lose their intended spatial extent. To address these issues, one path to explore is to add an optional polygon or free hand selection tool that can be used to define the extent taken into account by the spatial filter (Roberts, 2008). However, this has to be done carefully not to complicate the default interface as it will likely be more useful for expert users.

5.4 Clustering

Due to the freedom inherent to the process of VGI generation, data can accumulate and overlap in some areas. Data clutter can significantly hamper the visual exploration of VGI while at the same time they can provide useful information on the spatial repartition of the data collected. A distance-based clustering approach was implemented in an attempt to mitigate these issues (see Section 3.3.5.4). Participants were first guided through a series of steps to get familiar with the clustering tool. They had then two complete tasks of identification of areas of high concentration of VGI. Based on these activities, the participants were asked to provide feedback on the clustering tool in terms of improving map legibility and identification of areas with high concentration of drawings. It was expected that participants would be able to interpret the proportional circle relatively easily and its symbology and therefore would consider it as an important help for data browsing and map reading. Concerning the identification of drawing concentrations, the hypothesis was that the users would have some difficulties to quantify the number of drawings with the regular map display and thus the clustering function could bring them some insights on the spatial repartition of the drawings. However, it was expected that the simplified representation of the clusters could lead to some misinterpretations and questions on the accuracy and legitimacy of the clusters identified by the application.

5.4.1 Map Legibility

The participants were guided through a series of steps to experiment with the clustering feature and its related aspects such as the various symbolizations available and the adjustable clustering distance. Seventeen participants answered that the map was easier or much easier to read with the clustering activated (see Figure 5.4). These participants explained that the clusters simplify the map by limiting the number of overlapping features and clutter, and, as a result provide an overview of the comments' distribution. This confirms the hypothesis that users can be overwhelmed by the original drawings and their numerous shapes that are hard to separate out as their colours and contours blend together hindering the map reading. The aggregation of the drawings in a small number of proportional circles reduce the density of information on the map and isolate distinct symbols from the background which makes the map easier to process for the reader.

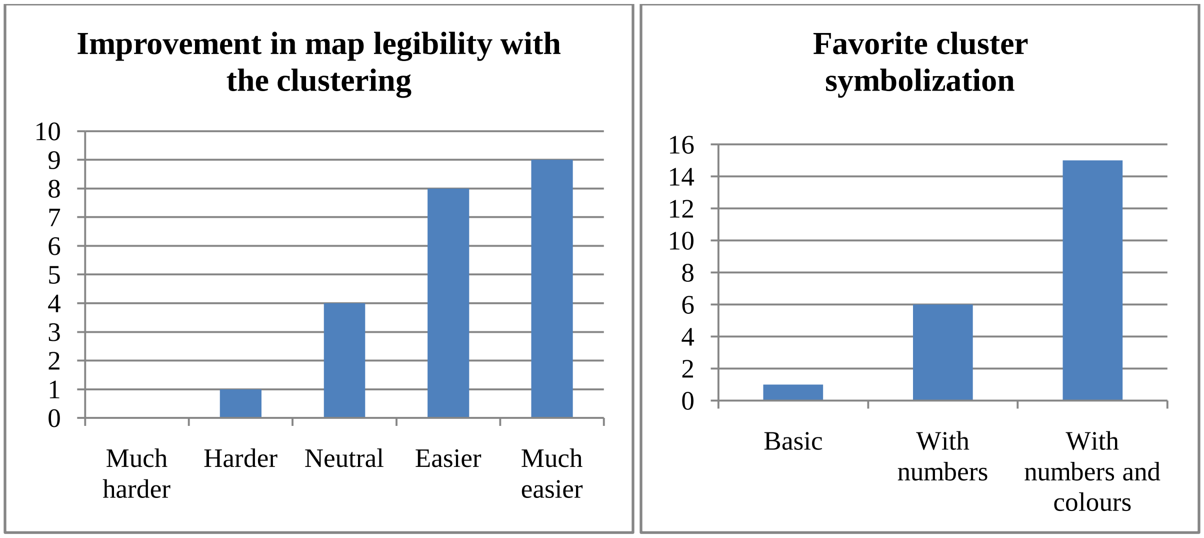


Figure 5.4: Map legibility with the clustering feature

However, four neutral ratings were also given by participants. These participants commented that each representation has its own advantage. The clustering gives a good sense of density and the number of drawings but it lacks important information such as the types of features (points, lines, and polygons), direction and coverage which can be found in the regular representation. One participant even judged the map harder to read with the clustering function as he/she did not like that the abstract cluster representation hides the exact nature and position of the features. This confirms the fact that individuals can possess different "global cognitive styles" that can influence their map reading and understanding as reported by MacEachren (2004 p. 202). Consequently, they influence the impact of the visualization tools on the individuals and as such they have to be taken into consideration at the time of design (MacEachren, 2004).

These cases underline the fact that it is essential to allow the users to switch quickly from one representation to another to enable them to explore the various aspects of the VGI, facilitate the understanding of the clustering representation and accommodate the users' preferences. It was possible in the application to switch the clustering function on and off which was satisfactory. However some enhancement to this capability can be done as suggested later in this section.

One of the main aspects of the participants' understanding of the clustering features is the symbolization of the clusters. Participants were able to switch between three types of symbolization, one using only the size to represent the number of drawings aggregated in a cluster, one using the size and a label indicating the number of drawings, and one using a colour scheme in addition to the size and the label (see Figure 3.11). Fifteen participants preferred the symbology that was based on a combination of circle colour, size and a label (see Figure 5.4).

Despite the fact that one participant mentioned that the circle sizes were very easy to visualize, a majority of participants preferred to have the two visual variables to portray a single attribute (i.e. the number of features clustered). This aligns well with Slocum et al. (2008) recommendation to use redundant symbolization for geovisualization tools as well as with MacEachren suggestion (2004, p. 88) that “combination should be useful in univariate map applications where the goal is to enhance discrimination while reinforcing appearance of order for quantitative information”. These suggestions are also supported by Dobson's (1983) experiment that showed that using the combination of variables (size and color value) improves the time and accuracy of the information processing over using size alone. Moreover, the participants also preferred having labels that specifies the number of drawings. This confirms that using anchors such as legend circles or labels is an efficient technique to help the user judge the relative magnitude of the symbols in addition to the perceptual scaling methods which are questioned by practitioners (MacEachren, 2004). Overall, it appears that it is recommended to use redundant symbolization along with anchors to facilitate the exploration of VGI.

Even though seventeen people rated the clustering feature as simple or very simple, four people rated the feature as complicated and facilitators observed some misinterpretations from some participants (see Figure 5.5). Two users intuitively thought that the circle sizes were related to the clustering distance or, in other words, that the symbol area represented where the drawings are located. However, the two are not related and the centres of the clustered features are not necessarily included in the circle representing the cluster or vice versa, drawings within the circle might not be part of the cluster (see Figure 5.6). For instance numerous features could be located in a very small area and generate a large symbol whereas a few features apart but still within the clustering distance generate a small symbol that does not cover them. This effect is even worsened by the fact that for lines or polygons, only their centres is taken account for the clustering calculation and therefore the drawing that is included in a cluster can span far outside the cluster symbol. These issues were also reported by two participants’ qualitative answers.

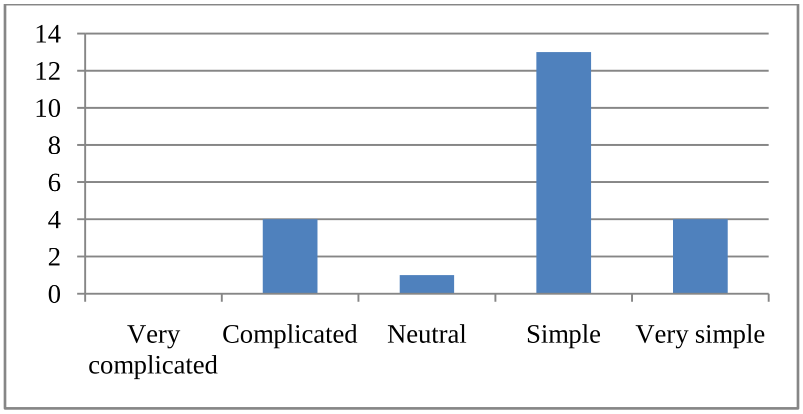


Figure 5.5: Reported complexity of the clustering function

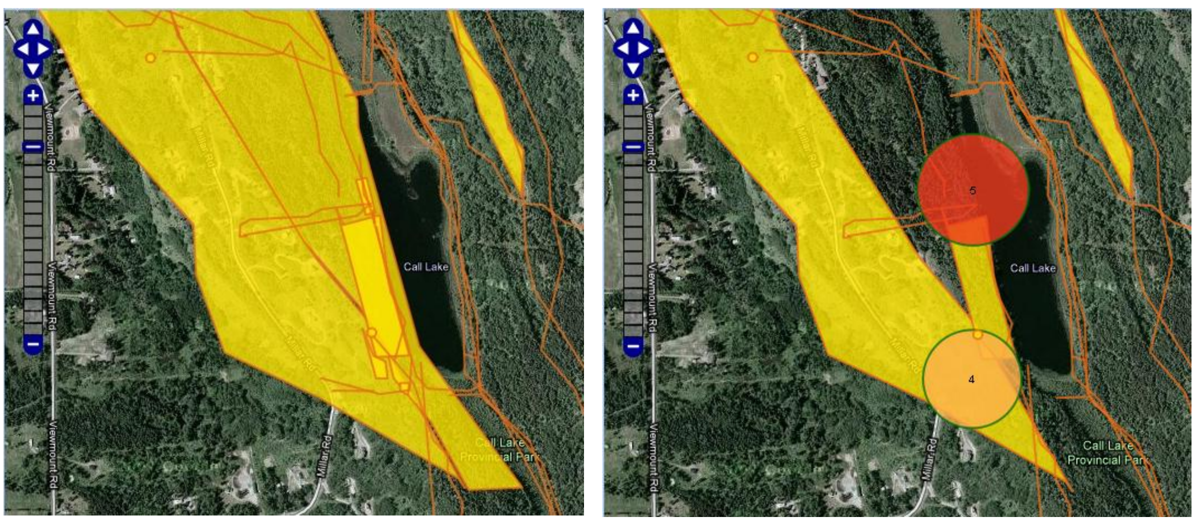


Figure 5.6: Example of misleading cluster representation

There are a few approaches that could alleviate these issues. First, allowing users to easily reveal the individual drawings contained within a given cluster through simple mouse over events could aid user comprehension. Second, a simple representation of the clustering distance directly on the map would greatly enhance the users' understanding as users cannot currently picture what the clustering distance actually represents on the map except by performing calculation with the scale bar.

Another solution could be to implement other clustering methods where the clustering distance is directly represented such as the grid-based approach. However this method has also serious limitations due to the modifiable unit area problem as mentioned in Section 3.3.5.4.

5.4.2 Identification of Areas of High Concentration of VGI

The participants were asked to complete two tasks relative to identifying areas of high concentration of VGI. In Part 3a and 3b of the workshop, they were asked to circle three areas they thought had an important concentration of user drawings using first the regular map display (manual method) and then the clustering function. As mentioned in Section 5.1, some participants forgot to save their drawings or got confused with the section numbers so their circling was not saved properly. Thus, the polygons of only twelve participants were taken into account for the results of this section (see Table 5.2). However, quantitative ratings and qualitative feedback of all the participants were considered as they were not flawed.

User	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Origin	UW									Bulkley Valley												
Level	Expert									N	Ex	N	Ex	N	N	Ex	N	Ex	Ex	Ex	Ex	Ex
Section 3a	4	3	3	3	1	3	3	2	1	3	3	3	3	3	3	1	1	0	3	3	3	1
Section 3b	3	3	3	0	3	4	3	0	0	3	3	5	0	0	3	0	1	4	3	3	3	3
Section 4	1	1	1	2	1	1	1	1	0	1	1	0	1	1	1	3	1	1	2	1	1	1
Section 5	3	3	3	3	3	2	3	0	1	1	3	3	0	8	0	4	3	3	3	3	2	1

Note: Ex= Expert; N= Novice

Table 5.2: Number of drawings saved by user and by part.

To study the participants' use of the clustering tool, the results obtained through each method were summarised by two raster maps presented in Figure 5.7. To generate the two rasters, the participants' polygons were first coded with a value of one inside their area. The polygons were then separated in a series of independent files with a file for each participant and each method so twelve files for each method were generated in total. These vector polygons were then rasterized to produce rasters with a value of one inside the outlined zones and zero outside. The raster of each participant were then added together for each method to finally produce two rasters with cell values representing the number of times the cell has been judged as an area of high concentration by the participants. Along with the summary rasters, calculations were performed on the vector polygons to present statistics on the polygons drawn by the participants (see Table 5.3).

	Manual	Clustering		Manual	Clustering	
User	**Sum of circled areas (sqkm)**	**Sum of circled areas (sqkm)**	**Difference between the two approaches (sqkm)**	**Number of assets contained**	**Number of assets contained**	**Difference between the two approaches**
1	48.96	10.21	**-38.75**	42	38	**-4**
2	25.44	28.83	3.40	63	57	**-6**
3	112.14	60.18	**-51.96**	90	76	**-14**
6	4.12	71.54	67.42	4	76	72
7	161.02	162.40	1.38	114	113	**-1**
10	33.64	6.76	**-26.89**	11	8	**-3**
11	51.71	213.95	162.24	38	118	80
12	93.19	64.76	**-28.43**	5	10	5
15	113.50	10.80	**-102.70**	28	15	**-13**
19	27.90	32.94	5.03	53	48	**-5**
20	63.47	46.34	**-17.13**	60	62	2
21	155.19	88.14	**-67.05**	101	84	**-17**
Total	890.29	796.84	**-93.45**	609	705	96
Average	74.19	66.40	**-7.79**	50.75	58.75	8.00

Table 5.3: Statistics about the identification of areas of high concentrations of VGI

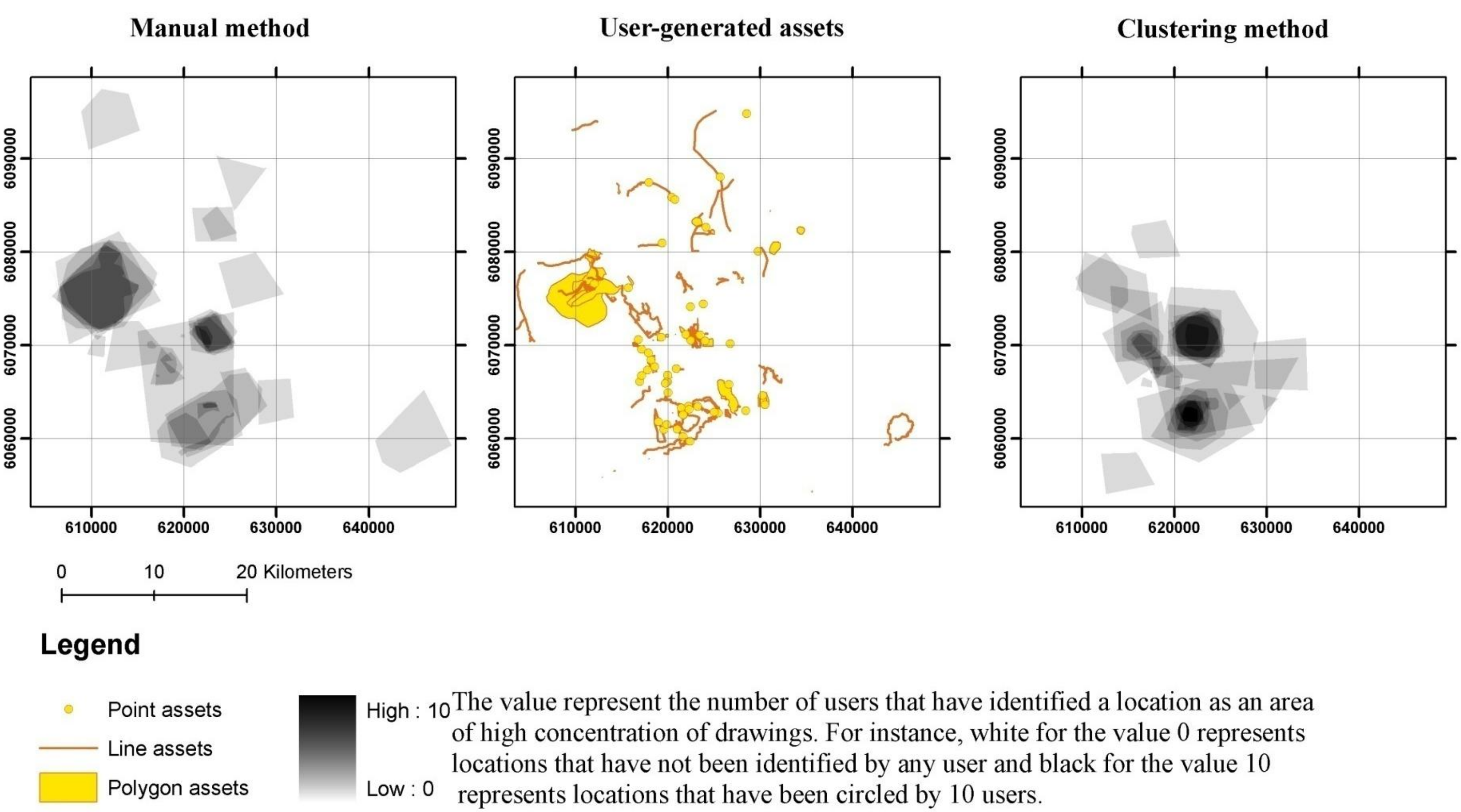


Figure 5.7: Comparison of the manual (Part 3a) and aided (Part 3b) methods

The results of the manual method in Table 5.3 show that participants adopt various behaviours when asked to circle areas that they think have a high concentration of VGI. The majority of participants use relatively large polygons and encircle an important number of drawings but some participants use small polygons and/or circle only a few features far from the actual areas of concentrations (see Table 5.3 and Figure 5.8). These two behaviours resemble the atomist and generalist profiles that McCleary (1975) identified in an experiment where participants delineated zones of different densities on a dot map (see Figure 5.9). According to McCleary, the generalists are able to grasp the overall pattern and draw only few simple lines whereas the atomists focus on details. Even though the reasons behind these individual differences are not clear and have yet to be investigated further, their potential influence on the use of the visualization functions and the meaning derived from VGI need to be considered (MacEachren, 2004). In our case, by comparing the results of "atomist like" participants such as user 6 and 11 (see Figure 5.8), it appears that the clustering functions might play a role in changing the strategy employed by the participants since both users switched from small areas to bigger areas. However, others like user 10 seem to be still focusing on details. While this is interesting, a larger number of participants is required to verify the impact of the visualization tools on the different cluster identification strategies.

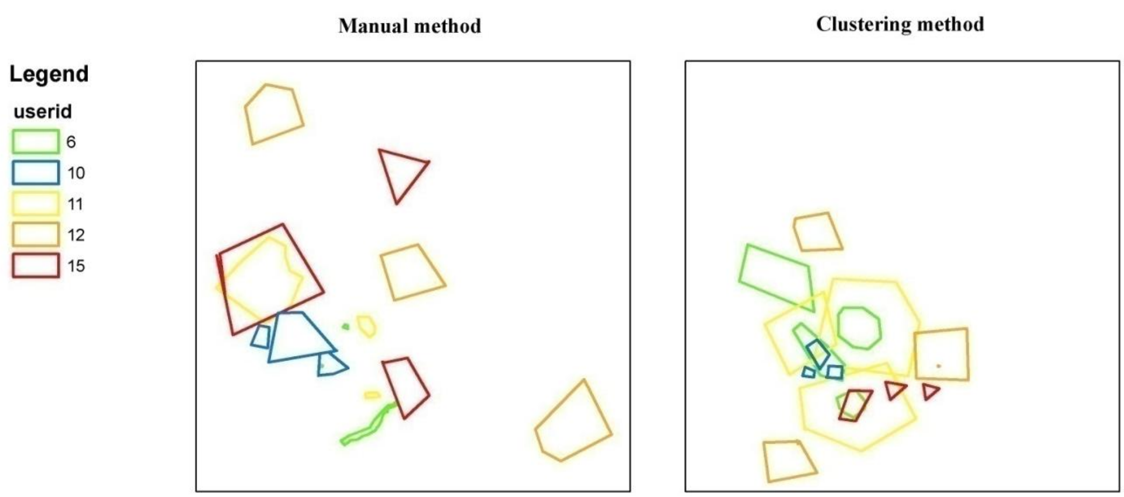
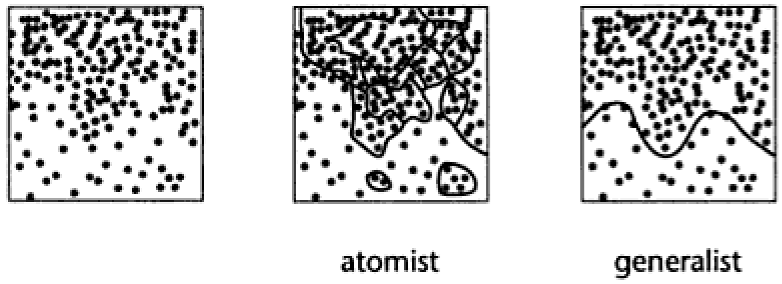


Figure 5.8: Areas identified by users focusing on small areas or a small number of features.



Source: adapted from McCleary (1975)

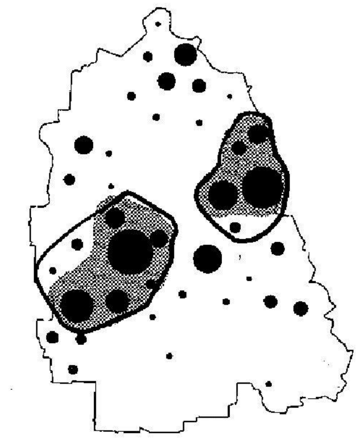
Figure 5.9: Example of the atomist and generalist profiles

The overall results presented in Figure 5.7 shows some significant differences in the areas of high VGI concentration identified with each method. Among these differences, one zone situated on left of the map brings some insights on the impact of the clustering function. This zone corresponds to a set of a few large polygon drawings overlapped by a few points and lines. It was circled by about half of the participants with the manual method. However, this zone does not have an especially high concentration of VGI. It seems that with the regular map display, the participants were not able to quantify the number of the drawings shown on the screen and relied more on the immediate impression provided by the map.

These results can be explained by referring to the theories of human visual perception and cognition that are involved in map reading in general and particularly in the task of identifying areas of high concentration of VGI. Three fundamental concepts of visual perception are particularly relevant in this case, namely perceptual grouping, figure-ground segregation and visual attention. Grouping refers to the way our vision define groups of elements from the sensory input. Based on the Gestalt psychology laws of grouping, MacEachren (2004) describes nine rules for perceptual grouping. The first two rules, proximity and similarity, are the most relevant here. Proximity refers to the fact that we tend to group objects that are close together. Similarity refers to the fact that we tend to group objects with similar characteristics such as colour and shapes. A second major element of our perceptual organization is the figure-ground segregation concept. It is summed up by MacEachren (2004, p. 107) as "figures that attract our attention are distinctive from the background and often appear to be in front of that ground". The figure-ground segregation process is characterised by six factors. Heterogeneity and contour are the two most relevant in our context. Heterogeneity refers to the difference between the figure and the background (e.g. a colour difference) and contour refers to the sharpness of the figure boundaries and directly relate to the heterogeneity. Finally, visual attention refers to the fact that our vision is selective and is directed to a location or an object with more or less precision comparable to a zoom lens (MacEachren, 2004). As a result, a broad initial view of a map captures only coarse features and then attention is progressively focused on particular features. The location of the attention is influenced by the properties of the map symbols. These three concepts (i.e. grouping, figure-ground and visual attention) interact with each other in complex ways to produce a representation from the perception of the map. Higher levels of cognition also intervene during map reading (e.g. mobilisation of knowledge and long-term memory), but in our case these aspects are not considered in details as they appear to play a lesser role as suggested by MacEachren (2004, p. 71): "From a cartographic perspective, low-level issues of grouping seem most relevant for exploratory cartographic visualization in which limited attention will be directed to any one map view and the goal is to notice patterns and relationship".

Building upon these principles, it appears that the drawings on the left side of the map (see Figure 5.7) attracted the visual attention of the participants because of a better figure-ground segregation due to the presence of larger polygons that contrast well with the imagery background. This initial perception obscured or dominated participants' assessment of feature counts in their identification of areas of concentration. Improving the symbology and the background data could mitigate this issue in some cases but it would not be sufficient overall, especially in dense areas.

The clustering function helped the participants to compensate the potentially misleading first interpretation by giving them another view of the data which focuses on the number of features rather than the area that each drawing covers. As a consequence, in the second task only two people circled the zone on the left of the map since it was not particularly stressed by the clustering function due to the low number of drawings. Instead, their focus was placed onto three other zones that had received less attention with the first method. These three zones have an actual concentration of VGI and as such were emphasized by the clustering function by circles of varying sizes. Therefore most of the participants identified these zone as areas of high concentration of VGI which results in two black and one dark grey areas in the Figure 5.7 that correspond to the areas outlined by the participants around some of the major circles or groups of circles generated by the clustering function. In this second task, the grouping of circles was mainly formed based on their proximity and similarity accordingly to the perception principles mentioned above and to the results of Slocum (1983) experiment. He showed that proximity, similarity and figure-ground segregation are the three main factors in the identification of clusters on graduated point maps (see Figure 5.10).



Note: in gray, the grouping predicted by Slocum; outlined in black the grouping done by the participant
Source: Slocum (1983)

Figure 5.10: Results from Slocum's experiment

These findings were confirmed by the subjective ratings of the participants. Eighteen people thought that it is easier or much easier to determine areas of concentration with the aid of the clustering function (see Figure 5.11). These ratings are further confirmed by the participants' qualitative feedback which indicated that it was easier and quicker to use the clustering function because it limits the guesswork that it is necessary with the regular display due to the various sizes and types of VGI.

Even though the results show that the clustering function gives the users new insights on the data, they also exhibit some of its limitations in term of accuracy and effectiveness. Indeed, since the clustering function helped the users to focus their attention on areas of high concentration of VGI, it is expected that the users should perform better at identifying them so the polygon they drew in the second task should encompass more VGI elements than in the first task. However, Table 5.3 shows that, except for two notable cases related to the atomist behaviour mentioned earlier (user 6 and 11), the number of assets contained is similar or even smaller than with the manual method. Several reasons can explained this unexpected result.

Participants focused on the cluster representation and mainly encircled the point symbol of the cluster. However as mentioned in the previous section, the point feature is a symbolization of the number of drawings within a cluster and it does not correspond to the actual area within which drawings are aggregated. This underlines again the fact that the link between the real features and the clusters need to be reinforced in the application to allow the participants to determine where the actual drawings are located. Furthermore, the location of the clusters can sometimes be approximate and confusing for the users. In part, this is due to the simple algorithm that uses the centre of the first aggregated drawing as a reference for the clustering distance and the location of the resulting cluster. As a consequence, the location can be misleading and is another reason for the equal or poorer performance with the clustering function. The limitations of the algorithm caused another point of confusion reported by one participant. Due to the fact that the drawings taken as references for creating clusters change at every map move, the clusters locations can change significantly when zooming in and out and it confuses the users. These issues coincide with the ratings on accuracy by the participants that include three and seven participants respectively rating the clustering as less or equally accurate (see Figure 5.11).

As mentioned in the previous section, these various issues can be mitigated by allowing the users to quickly switch from one representation to the other when moving the mouse over a cluster. This solution also facilitates the understanding of the clustering process and limits the black box effect that is felt by some people. Moreover, the default algorithm could be improved to make the representation more intuitive, for instance, a cluster can be located at the average centroid of all the aggregated drawings.

Participants also expressed some concerns about the meaning of the clusters. One person was concerned about the fact that the number of drawings would be assimilated to the "worth" of a place or assets while the density is not necessarily important while another participant mentioned that a cluster can be made up of features that are not necessarily related together in term of content. As mentioned in Section 3.3.5.4, these concerns are legitimate and call to better inform and educate the participants on how to derive information from the visualization to avoid misinterpretation. These concerns could also be mitigated by allowing enabling users to have more detailed cluster representations that take into account the number of replies about a drawing, the text filter entry, or the themes as it was started to be done with the pie chart representation that was implemented but not fully tested (see Figure 3.14).

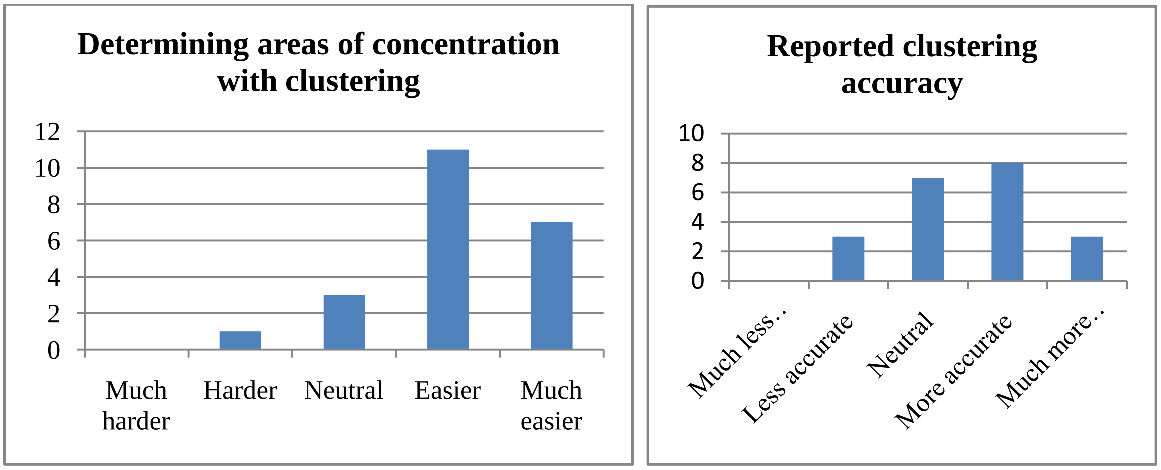


Figure 5.11**:** Participants ratings relative to the clustering features.

Overall, the clustering function helped the users in their exploration of VGI by giving them new insights on the data. However, the simple representation was sometimes too abstract or misleading and should be enhanced. These findings on the clustering feature should also be complemented by a study on the influence of the scale at which the users operate when they browse the data or identify clusters. Scale has important influence on the use of the clustering feature as it modifies the clustering distance and representation. This influence is confirmed by several users who mentioned that the clustering features was much more valuable with the clustering distance slider and by the software logs that show heavy use of the slider. However, the analysis of the distance slider and map movements data is out of the scope of this thesis. Finally, other ways of highlighting concentrations such as heat mapping (a term commonly used on the Internet to refer to a map overlay representing a 2D density estimate usually symbolized with a thermal-like colour scheme), that is becoming more and more common on the Internet and is starting to be applied to VGI, should be investigated and compared (White and Roth, 2010).

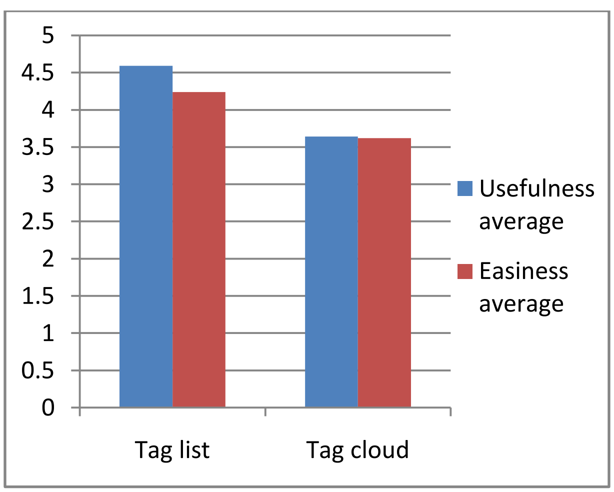
5.5 Tag-based Visualization

Building upon the tagging systems that enable the categorization of unstructured VGI, two tag visualization features were implemented in MapChat Viz. The tag cloud and tag list features permitted the users to view a summary of the tags associated with individual comments along with their respective number of occurrences. Combined with interactive linking and brushing, and spatial filtering techniques, they also enabled users to study the spatial distribution of the tagged comments (see Section 3.3.5.5). The main objective was to evaluate the merits and flaws of the tag visualization features for the exploration of VGI. To do so, participants were first asked to circle an area of their choice so each participant looked at a different part of the study area. They were then asked to report the tags of the area they picked ordered by number of occurrences. It was expected that the participants would value these features as they help them to deal with one of the main practical limitation of VGI, the lack of structure. Moreover, it was anticipated that despite the relative novelty of these concepts and visualization features, the participants would be able to complete the task quite easily based on their growing familiarity with the overall application and its underlying concepts such as linking and filtering. The secondary goal of this task was to determine whether the tag cloud was as useful and usable as its wide popularity amongst Web developers suggests or whether it was mainly a trendy design. As mentioned by Hearst and Rosner (2008, 2008), the actual purpose and usability of tag cloud are still being debated, thus it was expected that users would have mixed feelings about the tag cloud and would likely prefer the tag list because of its clearer and simpler design.

Most participants judged the theme visualization task as easy (11 people) or very easy (9 people) to complete, except two participants who rated it neutral. However, the users had different preferences as to which tool to use to complete the tasks. Overall, 15 users reported that they preferred to use the tag list over the tag cloud and they accordingly rated the tag list as easier to use and more useful (see Figure 5.12). The participants who favoured the tag list mentioned the fact that the ordering of the tags according to the number of occurrences and the explicit display of the number of occurrences were easier to understand and use than the alphabetical ordering and the variable font size of the tag cloud. These findings align well with the results of the few previous experiments on tag lists and tag clouds that have shown that users perform better and faster on tasks such as tag identification and impression formation (i.e. getting an overview of the data) with frequency ordered tag lists than with tag clouds (Halvey and Keane, 2007; Rivadeneira, Gruen, Muller, and Millen, 2007). Some of these participants, especially novices, were confused by the tag cloud visualization and questioned its usefulness. Five people gave a usefulness rating of two or less and three people mentioned that "the tag cloud was not necessary". These findings confirm the fact that lay users generally do not like the tag cloud visualization. They either find it messy or do not understand the meaning of the font size (Hearst and Rosner, 2008).

On the other hand, two users liked both visualization features mentioning that the list was good to get precise tag counts whereas the cloud was good to get an overall impression of the tag distribution and five users preferred the tag cloud with one participant even mentioning that "Tag clouds are great" (see Figure 5.12). Six out of these seven people were very familiar with GIS and the Internet and mostly young. This confirms the previous findings on the difference of responses to the tag cloud between expert and novice users. However, since the expert users were likely already familiar with tag cloud type visualizations, it also means that with a bit of practice the tag cloud can become an interesting alternative for information discovery tasks although they are limited when looking for specific information (Lohmann, Ziegler, and Tetzlaff, 2009; Sinclair and Cardew-Hall, 2008). Furthermore, as explained by Hearst (2008, p. 3), rather than being a good analysis tool, the tag cloud are a good indicator and conveyer of the feeling of the human activity by showing that "there are people actively using the information, commenting on it, and categorizing its contents".

The layout and visual design of a tag cloud have been proven to have an important influence on the user's performance (Lohmann et al., 2009; Rivadeneira et al., 2007; Seifert, Kump, Kienreich, Granitzer, and Granitzer, 2008). Therefore, the overall poor rating of the tag cloud may be attributable to some extent to its very basic implementation in MapChat Viz. The appearance and layout of the tag cloud can be improved significantly by following more best practices (Friedman, 2007; Lamentia, 2007a), improving the layout algorithm (Seifert et al., 2008) and adapting the layout to the various exploratory tasks (Halvey and Keane, 2007; Lohmann et al., 2009). For instance, a circular layout with decreasing popularity may have been more appropriate for the task given to the participants.



Note: left graph: 0 = not useful, 5 = useful; right graph: 0 = complicated, 5 = simple

Figure 5.12: Ratings of the tag list and tag cloud

To verify the participants' subjective evaluations and gain insights on the use of the features, the success of the participants at completing the task was evaluated by using the software logs. A reference tag summary was calculated for each user by determining the number of occurrences of each tag within each participant's area of interest. These reference tag summaries were then compared to the participants' answers (see Appendix A, Section 4) to evaluate how well the method performed. To ensure that the recorded areas of interest were valid, the calculations were performed only for the thirteen users who had saved correctly their work (see Table 5.2). Table 5.4 presents a sample of the data compiled for this discussion that is representative of the accuracy range of the participants' answers. Overall, the software logs confirm the positive participants' rating by revealing a good success of the participants at completing the task with no major error and five out thirteen users who completed the task very well reporting the most frequent tags in the right order. However, they also show some nuances as some inaccuracies in the participants' answers are evident with eight out of thirteen users who reported mostly the right tags but with some slight or important inaccuracy in the ordering. These results suggest that some limitations in both visualization features hindered the users' performance in the task.

User1		User2		User3		User5		User7		User10	
calculated order (# of occurrence)	**user answer**	**calculated order (# of occurrence)**	**user answer**	**calculated order (# of occurrence)**	**user answer**	**calculated order (# of occurrence)**	**user answer**	**calculated order (# of occurrence)**	**user answer**	**calculated order (# of occurrence)**	**user answer**
IR (6)	Tr	PN (6)	PN	Tr (19)	Tr	RF (3)	LF	PN (26)	RF	LF (7)	LF
Tr (5)	LF	LF (6)	RF	LF (11)	Ac	FF (3)	Ac	LF (20)	Tr	RF (5)	PN
RF (5)	IR	RF (4)	IR	Ac (10)	LF	LF (2)	Tr	RF (17)	LF	PN (4)	RF
Ac (5)	Ac	IR (2)	Tr	FF (8)	FF	EF (2)	RF	EF (8)	PN	EF (3)	
FF (5)	RF	Tr (1)	LF	IR (7)	RF	Ac (2)	EF	IR (6)	EF	IR (2)	
AN (3)	FF			RF (7)	IR	Tr (1)		Tr (5)	Ac	Tr(1)	
EF (2)				AN (4)				Re (3)	IR		
LF (2)				EF (3)				Ac (1)			
HRF (1)				HRF (1)							
				PN (1)							
Correct tags with small inaccuracies in the ordering		Correct tags with small inaccuracies in the ordering		Strong correspondence between actual and reported answer		Correct tags with small inaccuracies in the ordering		Correct tags with important inaccuracies in the ordering		Strong correspondence between actual and reported answer	

Note: *Ac = access; AN = access needed; EF = environmental feature; FF = future feature; HRF = historical recreational feature; IR = issue raised; LF = landscape feature; PN = protection needed; Re = resource; RF = recreational feature*

Table 5.4: Sample of the theme reviewing task results

One explanation for these results is to be found in the spatial interactive filtering element of the visualization functions. As with the spatial filter of the comment tab, the tag cloud and tag list contents were continually refreshed according to the drawings contained in the extent of the map being viewed. While this basic spatial filtering can produce a tag summary for a limited area quickly, it can be difficult to use effectively if the user focuses on non-regular areas. Some participants who defined complex areas of interest in the theme reviewing task reported that "It would be nice to create a cloud and list for a selected area" or "I could not find themes for just my polygon". To address this issue, the users need a manual way (i.e. freehand lasso or polygon) to select the area they want to filter the tag summary on (Roberts, 2008). Another more minor issue was caused by the fact that the tag summary spatial filter was constantly activated. Even though it was explained in the workshop script, this was not obvious for certain users who misinterpreted the visualization. As mentioned previously, meta-visualization techniques can help to solve such issues by giving visual clues on the status of the visualization tools to the users.

Like with the comment tab, the brushing and linking could be enhanced by supplementing the transient brushing on the tag names with a durable brushing. For instance, some users wanted to open pop-ups on features with a specific tag but when moving the mouse from the tag list to the feature the highlight disappeared. Furthermore, the durable brushing can permit to link the three views together, namely, comment tab, tag visualization and map, instead of two at a time. In this way, when a tag is selected, the relevant drawings are highlighted on the map and the content of the comment tab is accordingly filtered to display the comments on which the selected tag has been applied. Linking all the views together accordingly to multiple coordinated views principles would provide the users with more insights on the data and a more consistent user experience.

Fundamental in mapping, the method of grouping spatial features in separated thematic layers that are symbolized differently cannot be applied readily to this type of unstructured VGI since each feature can be tagged with several themes in a high number of unique combinations. A possibility to adapt these classic concepts is to combine the linking and brushing with the attribution of a different colour to each tag, as suggested by a participant, and to use this colour to highlight the related drawings on the map instead of using only one colour. This way, each feature is then virtually contained in many layers (i.e. a layer could be created for each tag) that can be visualized temporarily. Adapting existing core cartographic concepts is particularly interesting because the users can build upon their existing knowledge to use and understand the new tool (MacEachren, 2004).

However, this idea needs to be investigated further because, depending on the methods used for tagging, the number of tags can potentially be high and the associated colours becomes hard to distinguish as people can discriminate only about ten colours very well (MacEachren, 2004).To finish this section on linking, it is worth mentioning that other types of visualizations such as tag maps (i.e. tag cloud anchored directly on the map) can be interested to link the tags to the maps and need to be further investigated as done by Slingsby, Dykes, Wood, and Clarke (2007).

As with regards to the clustering, some participants expressed concerns about the fact that the frequency ordered list or the tag cloud highlight quantity and frequency and as a consequence seem to amalgamate these measures to the more subjective concept of importance. These fears align with some of the criticisms on the tag cloud reported by Hearst and Rosner (2008, p. 160) where the tag cloud is considered as “a bias towards popular ideas and the downgrading of alternative views”. As mentioned above, determining popular locations or themes is important but it is only one of the many possible perspectives on the data and users have to keep in mind that this apparent popularity is likely biased by the characteristics of the people volunteering the information, such as their interest, the places they frequent, and their access to the technology.

As mentioned in Section 4.3.2, a free tagging process was simulated to apply tags to the VGI to be able to experiment with the tag visualization features. Since the amount of VGI was relatively small and the tagging was simulated, the resulting tags did not present too many limitations, however in a real use case numerous limitations occur due to the inconsistent use of the tags by the different users (ambiguity: same words are used for two different things, synonyms: different words are used to describe the same things, etc..) (Mathes, 2004). Besides changing the methods of input for the tags to use a more rigid system with imposed tags or semi-rigid with suggested tags, the tag visualization can be enhanced to work around these issues and take advantage of the tagging richness. Some solutions to investigate are the use of tag clustering or hierarchical tag clouds amongst other methods (Lamentia, 2006). The tag clustering method that has already been proven useful consists in analyzing the groupings of tags assigned to the data and identifying common clusters of tags. As a result, it can reduce ambiguity and may promote a better understanding of the data among users as demonstrated in the Flickr tag search dialog (see Figure 5.13).

Note: the searched tag is "jaguar"

Figure 5.13: Tag clustering on Flickr

Overall, the frequency ordered tag list was found by the participants to be a simpler and clearer way to visualize a summary of the tags applied to the VGI features. On the other hand, the tag cloud need to be investigated further as novice users and even some expert users did not understand the particular implementation in MapChat Viz. It is not clear if the tag cloud would perform better with a more advanced implementation, a different context or perhaps, later in time should the concept becomes familiar among lay users (Lamentia, 2007b). Furthermore, both tag visualization variants can gain significantly in usability and usefulness by enhancing some aspects of the interactive filtering and linking. Finally, the process of data categorization through the tagging system is essential and is by itself a major enhancement that deals with the issue of the VGI structure mentioned in Chapter 2. One of the participants from the Bulkley Valley even commented in the final open-ended question "How much of my enjoyment of it is the tool, and how much is relief that someone finally, simplified the data?".

5.6 All Tools Enabled

As mentioned in Section 4.4, the last task of the workshop was less structured and permitted the participants to use the software in a more open-ended context, to determine if they would be able to take advantage of the tools and understand their relevance outside of a highly scripted environment. A secondary goal was to compare the usefulness of the features in a data exploration context. This task required participants to pick a theme of their choice (e.g. wildlife conservation, outdoor recreation...) and then to identify areas related to this theme that they would prioritize for special attention in the community. It was expected that most users would only use a subset of the simplest tools as they did not have enough time during the workshop to develop enough familiarity with the software to know when specific tools would be most appropriate. To support the results of this section, the software logs (see Section 3.3.5.6) were used to compile the total numbers of each type of action performed by each user during the task (see Table 5.5). Given the errors made by the users when saving their polygons, the log data of nineteen users out of twenty two was valid and used for the calculation (see Table 5.2). The numbers of events must be interpreted with caution as they represent various types of action. Some events represent instant actions such as map moves, mouse over a comment or over tag whereas others represent lasting actions such as the activation or deactivation of a feature like the clustering or spatial filter (see Table 3.1). With the former, the quantity of events is important to show the use whereas with the latter a count of one event can mean that the feature was activated during the whole task.

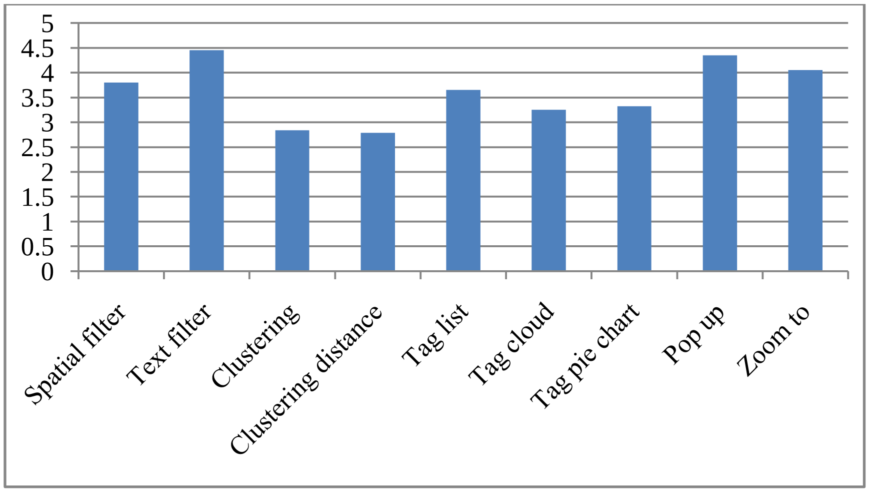
The objective of creating a task where the participants are able to capitalize on what they learned during the workshop did not work as well as expected. Indeed, many participants did not have much time to complete given that they had already devoted considerable amount of time completing the first four sections of the workshop script. This is confirmed by the software logs that show that eight out of nineteen participants only performed a very limited number of actions and used only a limited number of features (see the user 1, 4, 5, 6, 11, 12, 16, and 21 in Table 5.5).

Notwithstanding the real concerns of time limitations and possible user fatigue, the data collected in the section of the workshop do provide some insights into how participants valued and possibly understood the tools. While a few participants used almost all of the tools like user 3 and 7, most used only a limited number of features (see Table 5.5). Similarly, the quantitative ratings (see Figure 5.14) show that most participants found the basic features such as the text filter and the pop-up tool to be the most useful.

User / Event	1	2	3	4	5	6	7	9	10	11	12	15	16	17	18	19	20	21	22	Mean	Total by Event
Map moves	2	68	138	7	17	16	15	38	76	13	6	25	28	16	43	28	55	12	49	34.32	652
Mouse over comment	8	150	171	0	16	0	43	0	0	28	0	0	0	0	41	105	39	0	33	33.37	634
Zoom To	0	8	21	0	15	0	2	0	0	0	0	0	0	0	2	3	0	0	8	3.11	59
Popup	1	14	56	1	0	3	11	2	2	2	1	1	5	0	5	9	2	0	4	6.26	119
Text filter	1	4	25	0	0	1	7	1	0	5	0	0	0	0	1	4	2	3	3	3.00	57
Spatial filter	0	1	4	0	0	1	3	2	0	0	2	0	0	0	1	4	3	0	3	1.26	24
Clustering	0	0	10	1	1	4	5	2	0	1	0	1	1	1	0	1	3	1	3	1.84	35
Slider	0	0	1	0	0	0	1	6	0	0	0	0	0	0	0	0	0	0	1	0.47	9
Popup Paging	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.53	10
Mouse over tag	23	16	212	13	12	23	67	16	17	35	20	43	10	164	6	45	0	14	0	38.74	736
Tab change	3	6	49	1	6	7	30	17	1	7	1	5	5	3	5	9	12	5	10	9.58	182
Total by user	43	270	703	23	67	59	188	87	96	93	30	79	49	198	104	208	117	35	116	135	2565

Table 5.5: Software logs compiled for the section 5 of the workshop

Beyond the fatigue aspect mentioned above, these results suggest that either some features were actually not seen to be useful by the participants or they were not able to determine when a tool was relevant to use in an unscripted setting. Given the fact that the most users deemed the features useful in the previous sections when they were guided, the latter hypothesis is more likely to be correct. Moreover, it aligns with previous findings that mention that learning how to use a tool, its purpose and when it is useful are three different levels of understanding (Andrienko et al., 2002; Harrower, MacEachren, and Griffin, 2000). Users will often need some training and practice to be able to derive the full benefit from the tools. However, in the context of the Internet without the physical interaction of an in person workshop, training is difficult to provide. Online tutorials or interactive help within the tools are solutions that would need to be investigated over static help documentation (Andrienko et al., 2002). The workshop was concluded on a good note because participants left many enthusiastic remarks on the workshop and the application in the last open-ended question of the workshop such as “I love the visual upgrades, it is much more user friendly” or “Very interesting. Good planning tool. Thank you very much.”



Note: 0 = not useful at all; 5= very useful

Figure 5.14: Reported usefulness of the features to achieve Section 5 task

5.7 Summary of the Results

A series of features were tested during the workshop with the objective to evaluate the relative merits of the tools in terms of usability, usefulness and user satisfaction. Beyond testing the specifics of the implementation, the aim was to derive conclusions on the concepts underlying the design of the tools with regards to addressing the issues of browsing and filtering, spatially heterogeneous, unstructured quantitative data in VGI exploration and visualization.

Overall, the results of the usability studies can be seen as positive as the users gave very good quantitative ratings to the various visualization features in terms of ease of use and usefulness, accompanied by positive and enthusiastic qualitative remarks and suggestions. However, the facilitators' observations and the software logs have shown that the subjective ratings were sometimes slightly biased and that some improvement or extra features were needed. Nevertheless, the suggested enhancements should be implemented with special care to balance complexity and functionality. This can be achieved by using development methods that allows for basic and advanced interactions within one simple interface without adding a burden to novice users (Nielsen, 1993) and meta-visualization techniques to keep the users informed of the characteristics of the visualization environment (Roberts, 2008).

From the results of the usability studies, some conclusions on the ideas underlying the tools can be derived. The positive results on the comment tab, spatial filter and text filter and throughout the workshop show that the browsing and filtering of VGI was significantly enhanced by the multiple coordinated view practices implemented through interactive filtering, brushing and linking. While being careful to keep the tools at reach of novice users, these principles could be embraced even further by implementing full linking between all the views, more flexible filtering tool and durable brushing. The positive results on the clustering features have shown that the feature was helpful to give another perspective on the spatial repartition of the data to the users as well as to help them browse cluttered areas. However, the results have shown that it was not entirely appropriate for analysis purposes due to the extreme simplification of the representation. While the implementation could be improved by enhancing the cluster calculation and representation before further testing, it may be more efficient to keep the spatial clustering tool for browsing purposes and investigate alternative spatial density visualization techniques for analysis purposes such as heat maps that are becoming increasingly common on the Internet. Finally, even though the tagging process was only simulated, the resulting categorization of the data linked to the map appeared to aid users' exploration of VGI. The relatively negative answers about the tag cloud should not be seen definitive as users may get more familiar with the tool and the implementation in MapChat Viz could be improved and then further tested. Overall, despite some limitations in the data, these results show the potential of the geovisualization techniques to enhance VGI exploration.

Chapter 6
Conclusion

The overall goal of this thesis was to explore the potential offered by visualization techniques to deal with the challenges related to the utilization of VGI. This effort represents a first step towards taking better advantage of VGI whose potential is still underexploited. The first section outlines the contributions that were made to the VGI research through the completion of the thesis objectives. Some of the project limitations are discussed, highlighting suggestions to improve future studies. This chapter concludes with recommendations for possible research avenues, stemming from the work and results presented in this thesis.

6.1 Summary and Contributions

The first objective was to establish a VGI typology based on the review of the literature on the Geospatial Web, PPGIS and VGI. The VGI research field is still relatively young and no other similar VGI frameworks or typologies had been developed to date. Only few suggestions of ways to differentiate types of VGI were mentioned throughout the literature. Thus, with no prior complete framework to build upon, developing a VGI typology was necessary to provide a foundation to the thesis. Based on the existing body of literature, a comprehensive typology was elaborated. The typology defines a continuum that spans the different types of VGI. This continuum is anchored by three main types that are identified and characterized by a set of properties, namely scientific knowledge (SK), local knowledge (LK), and personal knowledge (PK) also referred to respectively as VGI type 1,2, and 3. This typology was then used to frame the work presented in this thesis that focused on Type 2 VGI. Beyond being a foundation for this thesis, the typology is an important contribution as it fills a gap in the literature and gives an overview and a structure to the recently emerged VGI research field that is still unstructured, teeming with new works and ideas. While this typology is only one of several possible approaches to characterize the different types of VGI, it provides a structure to explore differences and synergies between the various types VGI and it can serve as a general base to organize future works on VGI.

The second objective focused on reviewing the challenges related to the use and visualization of VGI and their potential solutions. As mentioned along the thesis, VGI constitutes an unprecedented source of data due to its distributed creation process and its unique characteristics. As such, it has the power to complement and maybe, on some occasions, supersede regular and official sources of data. Type 2 VGI is especially rich and complex as it conveys types of information that were rarely mapped in the past such as qualitative, unstructured and subjective data. However, these characteristics also raise a number of challenges that hamper its utilization and need to be tackled to fully benefit from the power of VGI. To fulfil the second objective, a series of general challenges were identified based on the literature on VGI and the unique characteristics of VGI defined in the typology. This discussion was complemented by a review of the different solutions that have been investigated to date. Building upon this broad review, the thesis focused on the challenges related to information overload and extraction of meaning, and on the solutions using interactive geovisualization techniques rather than automated and processing-based approaches. Generic needs for the visualization of Type 2 VGI with respect to these challenges were identified and generic solutions based on geovisualization techniques and Internet-based tools were discussed. One main contribution can be emphasized in this work. Even though the use of Web-based geovisualization techniques have been formalized in the general context (Harrower, Robinson, Roth, and Sheesley, 2009), their use with VGI has to date been limited to specific implementations spread over the Internet. Thus, the formalization of the use of geovisualization techniques with respect to generic VGI visualization needs is a methodological contribution that provides general guidelines to design future Web-based VGI software.

The third objective concerned the design and development of a prototype Web tool that implements some of the generic visualization techniques. The underlying goal was to provide an environment that brings together several visualization techniques in order to evaluate their merits. A secondary requirement was to make the software accessible to community-based organizations that are likely to take advantage of Type 2 VGI as discussed in Section 2.1 about PPGIS. This objective was met by developing a prototype with a set of open source components. A number of techniques ranging from basic browsing to dynamic spatial clustering and theme visualization were implemented. Beyond the specific implementation, the completion of this objective resulted in the contribution of a software design approach that enables the formal testing and comparison of visualization techniques. First, it brings together several visualization techniques that were dispersed across the Internet. Second, it offers the possibility to use the software in a step-wise fashion to progressively reveal the different visualization tools in the interface. As a result, it provides a testing environment that limits confounding factors, offers a gradual learning curve for the participants, and can be easily coupled with a workshop script and questionnaire to collect feedback. Few examples of such an approach can be found in the usability engineering literature. However, some approximations were made in the implementation which resulted in some limitations in the data collected. These issues and related suggestions for improvement are further discussed in the next section. Besides these methodological contributions, small contributions were made to the documentation and code bases of open source projects and more will be made as the code implemented matures. The successful development of this prototype also testifies of the maturity and power of the OSGIS projects that permitted to accommodate the time and resource constraints. They enabled a quick and flexible software development by providing polished out of the box features along with the possibility to access the source code to develop and integrate customised tools. Additional to the fact that software had no licensing fees, a qualified and responsive gratuitous support was received from the open source community provided that issues were clearly formulated and investigated thoroughly before seeking help.

The fourth objective was to test the prototype tool with a cross-section of potential users to assess the relative merits of the selected techniques. To complete the objective, it was necessary to collect a VGI dataset that is an adequate example of Type 2 VGI and its related challenges. This was achieved by building upon previous research work involving citizen-generated data collection and the MapChat software. Even though not a direct contribution of this thesis, this prior research work is an example of a successful collection of Type 2 VGI that was initially nurtured by university-based researchers and that was then adopted and maintained by the local population for their own interests. Such examples are rare in the PPGIS literature as the initiatives usually wind down with no real outcome for the local population. In order to test the prototype with the user-generated asset data, human-computer interaction and usability engineering methods were employed. As such, this study is a successful example of light-weight usability testing that can be achieved with relatively little resources and should therefore be more systemically pursued (Nielsen, 1994). Another contribution of the testing performed in this thesis is that while most of these techniques existed in one form or another, they had rarely if ever been formally tested to prove their usability and utility.

As discussed in Chapter 5, the overall positive results of the techniques evaluation demonstrate the potential of visualization techniques to mitigate the challenges of information overload and meaning extraction inherent to the utilization of Type 2 VGI. These techniques enabled the users to quickly find information that is relevant spatially and/or in term of content and to gain new insights on this rich but complex data that they could not simply acquire with a regular display. Thus, while many VGI projects halt at the contribution step, these results show that simple visualization techniques are one way to take it further, which can be implemented and integrated with existing projects without significant efforts. Beyond confirming that user with previous experience with GIS and Internet are savvier with such techniques (Meng and Malczewski, 2009), this case study has also shown that users with little computer and Internet experience are capable of taking advantages of geovisualization techniques.

However, even though these techniques are easy to use, it was stressed several times that it should not cause the developers and users to forget that any visualization techniques are inherently subjective; similarly to regular maps that represent subjective representations based on the cartographers' choices (Dodge et al., 2008b). Therefore, ways to convey critical spatial and visualization thinking (Goodchild, 2010) through online training or alternative visualization techniques need to be investigated.

6.2 Limitations

Although the research objectives were fulfilled with success, this thesis has several limitations at the research and software design levels. Related to the research design, three main limitations were identified. First, the recruitment of the participants was done mainly by email and with no specific incentive besides the personal interest of the individuals in the case study area and/or GIS technologies. This resulted in a relatively small representation of novice users which induced a bias in the evaluation results. To resolve this issue, more resources need to be allocated to the recruitment process and small financial compensations to the participants.

The overall number of participants (22 in total) was too small to perform statistical analyses but it was satisfactory to perform qualitative analysis of the feedback and software usage data collected (Nielsen, 1993). However, while all the questionnaires were properly completed by the participants, an error in the saving process of the participants' inputs within the software significantly reduced the amount of data available for the analysis of the results of two tasks (only 12 valid results) and consequently reduced the validity of the observations. A recommendation for future works using a similar approach for software testing is to include even more obvious messages in the software to inform people of their progression as well as to implement a strict control system to ensure that every user has drawn the appropriate number of shapes for each task.

Third, the analysis of the results of the evaluation was hampered by the fact that some users did not complete the tasks as expected. This was particularly true for the last task which was meant to be open-ended to evaluate the tools in a broader context but only resulted in limited uses of the software. As discussed in Section 4.4.3, these challenges are to some extent related to the fact that tasks are difficult to define when testing geovisualization tools, as they are designed to deal with ill-defined problems. However, in this case, it was also due the fact that a limited amount of time could be asked from the participants as they were participating on a volunteer basis. Thus, some compromises were made on the task lengths in an attempt to get the most possible feedback from the users. A recommendation for future works would be to focus on a smaller set of tasks either open-ended or scripted tasks to ensure the proper completion of the tasks.

Two main limitations pertaining to the software design are identified. First, in its current implementation, the software is limited as to the number of vector objects it can handle without becoming too slow to be usable. This limitation is due to the fact that all the vector data are loaded in the client browser to enhance the interactivity but, to date, Web browsers have limited capabilities to handle vector data. Until the necessary improvements are implemented in the browsers, for instance through the implementation of the upcoming HTML5 standards, some workarounds based on a combination of image and vector data can be investigated. An example of such an implementation can be found in MapChat 2 (Hall and Leahy, 2010).

Lastly, as the software was implemented as a prototype and proof of concept, it did not fully comply with the generic design guidelines described in Section 3.2.4 such as the full linking between all the views. This was mainly due to the fact that each tool was programmed individually. To develop a more complete visualization application, generic coding structure that manages the views and linkages should be used as a base for the development. This will result in an easier and more systematic management of the views and linkages. Examples of such approaches can be found in geovisualization tool kits such as Improvise (Weaver, 2004).

6.3 Recommendations for Future Research

As an emerging research field, positioned at the intersection of several domains, the VGI research has countless research avenues to explore. Several of them can be identified by building upon the work and results presented in this thesis. First, as discussed in Section 3.1, geovisualization can have several roles throughout the data exploratory process ranging from exploration and experimentation with the data to the synthesis and presentation of hypotheses and ideas. While this thesis focused on the first stages of this process, it is necessary to investigate techniques to enable the users to perform successfully the last steps of the process. Geovisualization methods, such as the possibility to store the history of the data handling, to explore different reasoning paths, or to save searches and specific representations, are examples of techniques that can be used to provide the users with tools to demonstrate and present information and ideas that they gained through the data exploration process. As a result, users are able to be empowered through the use of VGI as they can have access to a streamlined process from the unknown and complex VGI dataset to a set of information that is presentable and usable to advocate a case.

As discussed in Section 3.2.4, using geovisualization techniques with VGI is particularly powerful as it facilitates the exploration of the multiple VGI dimensions and their relationships. However, several facets of VGI that were not investigated in this thesis should be studied in future work. First, the dynamic aspect of VGI was not exploited. The long standing and recent literature on cartography and geovisualization abounds with time visualization techniques ranging from the simple time slide to more advanced animation and filtering (Harrower, 2009; Harrower and Fabrikant, 2008; Slocum, McMaster, Kessler, and Howard, 2008c) and numerous examples of Web-based implementations are rapidly appearing. Second, the three way interactions between users that are at origin of the VGI creation also need to be investigated. Such work could include the visualization of the social interaction through graph visualization or the relationships between different opinions and viewpoints such as the emergence of conflict or consensus.

While this thesis focused on interactive visualization approaches, two other major avenues to improve the utility of VGI were mentioned in Chapter 2, namely the use of the power of crowd collaboration with techniques such as folksonomy and the use of computing and algorithm-based methods with techniques such as computational linguistics. Investigating further these three avenues and even more importantly their potential combinations constitute the way forward to improve the utility of VGI for the society. The needs for such advanced techniques is becoming more and more critical as VGI is starting to be used for crisis response and for decision making by governments that have started to embrace the concepts of transparency, public participation and collaboration through the movement termed Government 2.0.

References

Al-Kodmany, K. (2001). Visualization Tools and Methods for Participatory Planning and Design. *Journal of Urban Technology, 8*(2), 1.

Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J., and Kretchmer, U. (2002). Testing the usability of interactive maps in CommonGIS. *Cartography and Geographic Information Science*, *29*(4), 325(18).

Arnstein, S. R. (1969). A Ladder Of Citizen Participation. *Journal of the American Planning Association, 35*(4), 216. doi:10.1080/01944366908977225

Aysegul, T., and Roche, S. (2007). Classification fonctionnelle des Public Participation GIS. *Revue Internationale de géomatique, 18*(4), 429-442.

Baeker, G. (2010). *Rediscovering the Wealth of Places: a municipal cultural planning handbook for Canadian communities*. Municipal World. Retrieved from http://www.municipalworld.com/index.php/BookShop/RediscoveringTheWealthOfPlaces

Balram, S., and Dragicevic, S. (2006). *Collaborative Geographic Information Systems*. Idea Group Publishing.

Bishr, M., and Mantelas, L. (2008). A trust and reputation model for filtering and classifying knowledge about urban growth. *GeoJournal*, *72*(3), 229-237. doi:10.1007/s10708-008-9182-4

Bojórquez-Tapia, L. A., Diaz-Mondragón, S., and Ezcurra, E. (2001). GIS-based approach for participatory decision making and land suitability assessment. *International Journal of Geographical Information Science*, *15*(2), 129-151.

Bradley, T. (2009, November 16). Google Street View Battle Highlights Privacy Challenge. *PCWorld through Reuters*. Retrieved from http://www.reuters.com/article/idUS236242272220091116

Bugs, G., Granell, C., Fonts, O., Huerta, J., and Painho, M. (2010). An assessment of Public Participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil. *Cities*, *27*(3), 172-181. doi:10.1016/j.cities.2009.11.008

Carver, S. (2003). The Future of Participatory Approaches Using Geographic Information: Developing a Research Agenda for the 21st Century. *URISA Journal*, *15*(APA 1), 61-71.

Carver, S., Evans, A., Kingston, R., and Turton, I. (2001). Public participation, GIS, and cyberdemocracy: evaluating on-line spatial decision support systems. *ENVIRONMENT AND PLANNING B-PLANNING & DESIGN*, *28*(6), 907-921.

Catt, D. (2008, August 8). Location, keeping it real on the streets, yo! *Code: Flickr Developer Blog*. Retrieved August 18, 2009, from http://code.flickr.com/blog/2008/08/08/location-keeping-it-real-on-the-streets-yo/

Chen, R., and Xie, J. (2008). Open Source Databases and Their Spatial Extensions. In *Open Source Approaches in Spatial Data Handling* (pp. 105-129). Retrieved from http://dx.doi.org/10.1007/978-3-540-74831-1_6

Chipeniuk, R. (2004). Planning for amenity migration in Canada - Current capacities of interior British Columbian mountain communities. *MOUNTAIN RESEARCH AND DEVELOPMENT*, *24*(2004), 327-335.

Chipeniuk, R. (2008). Some tools for planning for amenity migration in remote rural settlements: lessons from participatory action. *Community Development Journal*, *43*(2), 222-238.

Cowen, D. J. (2007). The Availability of Geographic Data: The Current Technical and Instituional Environment. In *The Handbook of Geographic Information Science*. Wiley-Blackwell.

Craglia, M. (2007). Volunteered Geographic Information and Spatial Data Infrastructures: when do
parallel lines converge? Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/participants.html

Craig, W. J., Harris, T. M., and Weiner, D. (2002a). *Community Participation and Geographical Information Systems* (1st ed.). CRC.

Craig, W. J., Harris, T. M., and Weiner, D. (2002b). Community Participation and Geographical Information Systems. In *Community Participation and Geographical Information Systems* (1st ed.). CRC.

Dana, P. H. (2007). Surveys of People and Place. In *The Handbook of Geographic Information Science*. Wiley-Blackwell.

Davies, C., Chao, L., and Albrecht, J. (2010). Human understanding of space. In *Interacting with Geospatial Technologies* (pp. 19-35). Wiley-Blackwell.

De Longueville, B., Ostländer, N., and Keskitalo, C. (2009). Addressing vagueness in Volunteered Geographic Information (VGI) –
A case study (under review). *International Journal of Spatial Data Infrastructures Research,*, *4*(11). Retrieved from http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/132/137

DiBiase, C. (1990). Visualization in the Earth Sciences at Penn State. *Earth and Mineral Sciences, Bulletin of the College of Earth and Mineral Sciences*, *59*(2), 13-18.

Dobson, M. W. (1983). Visual information processing and cartographic communication: the utility of redundant stimulus dimension. In *Graphic communication and design in contemporary cartography (Progress in contemporary cartography)* (pp. 149-175). Chichester [West Sussex] ;;New York: Wiley.

Dodge, M., McDerby, M., and Turner, M. (2008a). *Geographic Visualization: Concepts, Tools and Applications*. Wiley Blackwell.

Dodge, M., McDerby, M., and Turner, M. (2008b). The Power of Geographical Visualizations. In *Geographic Visualization: Concepts, Tools and Applications*. Wiley Blackwell.

Dörk, M., Carpendale, S., Collins, C., and Williamson, C. (2008). VisGets: coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1205-1212. doi:10.1109/TVCG.2008.175

Doyon, J., and McKenna, J. (2009). WFS Server — MapServer 5.4.0 documentation. Retrieved July 20, 2009, from http://mapserver.org/ogc/wfs_server.html

Dunn, C. (2007). Participatory GIS - a people's GIS? *Progress in Human Geography*, *31*(5), 616-637.

Dykes, J., MacEachren, A. M., and Kraak, M. (2005). *Exploring Geovisualization*. Pergamon.

Dykes, J., Purves, R., Edwardes, A., and Wood, J. (2008). Exploring Volunteered Geographic Information to Describe Place:
Visualization of the 'Geograph British Isles' Collection. In *Proceedings of the GIS Research UK 16th Annual Conference GISRUK 2008*. Presented at the GISRUK, Manchester.

Elwood, S. (2002). The impacts of GIS use for neighbourhood revitalization in Minneapolis (1st ed., pp. 77-88). CRC.

Elwood, S. (2008a). Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. *GeoJournal*, *72*(3), 133-135. doi:10.1007/s10708-008-9187-z

Elwood, S. (2008b). Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, *72*(3), 173-183. doi:10.1007/s10708-008-9186-0

Elwood, S. (2009). Geographic Information Science: New Geovisualization Technologies -- Emerging Questions and Linkages with GIScience Research. *Progress in Human Geography*, *33*(2), 256-263.

Elwood, S., and Ghose, R. (2001). PPGIS in community development planning: Framing the organizational context. *Cartographica*, *38*(3-4), 19-33.

Esnard, A. M. (2007). Institutional and Organizational Barriers to Effective Use of GIS by Community-Based Organizations. *URISA Journal*, *19*(2), 13-21.

Euvrard, R. (2009, April 8). A la découverte des Location Based Social Networks, les « néo-géo-réseaux ». *GeoInWeb*. Retrieved from http://www.geoinweb.com/

2009/04/08/location-based-social-networks/

Evans, A. J., Kingston, R., and Carver, S. (2004). Democratic input into the nuclear waste disposal problem: The influence of geographical data on decision making examined through a Web-based GIS. *Journal of Geographical Systems*, *6*(2), 117-132. doi:10.1007/s10109-004-0128-1

Fisher, P. F. (1999). Models of uncertainty in spatial data. In *Geographical Information Systems: Principles, Techniques, Applications and Management, 2nd Edition* (2nd ed.). Wiley.

Flanagin, A., and Metzger, M. (2008). The credibility of volunteered geographic information. *GeoJournal*, *72*(3), 137-148. doi:10.1007/s10708-008-9188-y

Friedland, G., and Sommer, R. (2010). Cybercasing the Joint: On the Privacy Implications of Geotagging. In *Proceedings of the Fifth USENIX Workshop on Hot Topics in Security*. Washington, DC, US. Retrieved from http://www.icsi.berkeley.edu/cgi-bin/pubs/publication.pl?ID=002932

Friedman, V. (2007). Tag Clouds Gallery: Examples And Good Practices. *Smashing Magazine*. Retrieved July 28, 2010, from http://www.smashingmagazine.com/2007/11/07/tag-clouds-gallery-examples-and-good-practices/#comments

Fuhrmann, S., Ahonen-Rainio, P., Edsall, R. M., Fabrikant, S. I., Koua, E. L., Tobón, C., Ware, C., et al. (2005). Making Useful and Useable Geovisualization: Design and Evaluation Issues. In *Exploring Geovisualization*. Pergamon.

Fuller, T., Guy, D., and Pletsch, C. (2002). *Asset Mapping: A Handbook*. Agriculture and Agrifood Canada. Retrieved from http://www.rwmc.uoguelph.ca/document.php?d=11

Ghose, R. (2001). Use of Information Technology for Community Empowerment: Transforming Geographic Information Systems into Community Information Systems. *Transactions in GIS*, *5*(2), 141-263. doi:10.1111/1467-9671.00073

Gibson, R., and Erle, S. (2006). *Google Maps Hacks*. O'Reilly Media, Inc.

Goodchild, M. (2002). Foreword. In *Community Participation and Geographical Information Systems* (1st ed.). CRC.

Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211-221. doi:10.1007/s10708-007-9111-y

Goodchild, M. (2008). Commentary: whither VGI? *GeoJournal*, *72*(3), 239-244. doi:10.1007/s10708-008-9190-4

Goodchild, M. (2010). *From Community Mapping to Critical Spatial Thinking*. Presented at the National Science Foundation (NSF) Distinguished Lecture Series. Retrieved from http://www.nsf.gov/news/news_summ.jsp?cntn_id=118063&org=NSF&preview=false

Gorman, S. (2006, July 20). Moving Past Push Pins. *Off the Map*. Retrieved from http://blog.fortiusone.com/2006/07/20/moving-past-push-pins/

Gouveia, C. (2004). Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management*, *71*(2), 135-154.

Gouveia, C., and Fonseca, A. (2008). New approaches to environmental monitoring: the use of ICT to explore volunteered geographic information. *GeoJournal*, *72*(3), 185-197. doi:10.1007/s10708-008-9183-3

Grenier, L. (1998). *Working with Indigenous Knowledge: A Guide for Researchers*. International Development Research Center Books.

Haklay, M. (2008). How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. *submitted to Environment and Planning B*.

Haklay, M. (2009). *Beyond good enough? Spatial Data Quality and OpenStreetMap data*. Presented at the State Of The Map, Amsterdam. Retrieved from http://www.slideshare.net/mukih/beyond-good-enough-spatial-data-quality-and-openstreetmap-data

Haklay, M., and Nivala, A. M. (2010). User-centred design. In *Interacting with Geospatial Technologies*. Wiley-Blackwell.

Haklay, M., and Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. *Pervasive Computing, IEEE*, *7*(4), 12-18.

Haklay, M., and Tobón, C. (2003). Usability evaluation and PPGIS: towards a user-centred design approach. *International Journal of Geographical Information Science*, *17*(6), 577. doi:10.1080/1365881031000114107

Haklay, M., Singleton, A., and Parker, C. (2008). Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, *2*(6), 2011-2039. doi:10.1111/j.1749-8198.2008.00167.x

Hall, G. B., Chipeniuk, R., Feick, R. D., Leahy, M. G., and Deparday, V. (2010). Community-based production of geographic information using open source software and Web 2.0. *International Journal of Geographical Information Systems*.

Hall, G. B., and Leahy, M. (2010). Map-chatting within the Geospatial Web. In *Advances in Web-based GIS, Mapping Services and Applications*. CRC Press.

Hall, G. B., and Leahy, M. G. (2008). Design and implementation of a map-centred synchronous collaboration tool using open source components: the MapChat project. In *Open Source Approaches in Spatial Data Handling*. Springer.

Hall, G. B., Moore, A., Knight, P., and Hankey, N. (2009). The extraction and utilization of local and scientific geospatial knowledge within the Bluff oyster fishery, New Zealand. *Journal of Environmental Management*, *90*(6), 2055-2070. doi:10.1016/j.jenvman.2007.08.022

Halvey, M. J., and Keane, M. T. (2007). An assessment of tag presentation techniques. In *Proceedings of the 16th international conference on World Wide Web* (pp. 1313-1314). Banff, Alberta, Canada: ACM. doi:10.1145/1242572.1242826

Harrower, M. (2009). Map Animation. In *Cartography 2.0*. Retrieved from http://cartography2.org/

Harrower, M., and Fabrikant, S. (2008). The Role of Map Animation for Geographic Visualization. In *Geographic Visualization: Concepts, Tools and Applications*. Wiley Blackwell.

Harrower, M., MacEachren, A. M., and Griffin, A. L. (2000). Developing a Geographic Visualization Tool to Support Earth Science Learning. *Cartography and Geographic Information Science*, *27*, 279-293. doi:10.1559/152304000783547759

Harrower, M., Robinson, A., Roth, R. E., and Sheesley, B. (2009). *Cartography 2.0*. Retrieved from http://cartography2.org/

Harrower, M., and Sheesley, B. (2005). Designing Better Map Interfaces: A Framework for Panning and Zooming. *Transactions in GIS*, *9*(2), 77-89.

Hearst, M. A. (2008). What's Up with Tag Clouds? *Perceptual Edge, Visual Business Intelligence Newsletter*. Retrieved from http://www.perceptualedge.com/articles/guests/whats_up_with_tag_clouds.pdf

Hearst, M. A., and Rosner, D. (2008). Tag Clouds: Data Analysis Tool or Social Signaller? In *Hawaii International Conference on System Sciences* (Vol. 0, p. 160). Los Alamitos, CA, USA: IEEE Computer Society. doi:http://doi.ieeecomputersociety.org/10.1109/HICSS.2008.422

Hersman, E. (2009, February 4). Crisis Info: Crowdsourcing the Filter. *The Ushahidi Blog*. Retrieved August 19, 2009, from http://blog.ushahidi.com/index.php/2009/02/04/crisis-info-crowdsourcing-the-filter/

Holck, J., Pedersen, M., and Larsen, M. (2005). Open source software acquisition: beyond the business case. *13th European Conference on Information Systems*.

Honan, M. (2009, January 19). I Am Here: One Man's Experiment With the Location-Aware Lifestyle. *Wired*. Retrieved from http://www.wired.com/gadgets/wireless/magazine/17-02/lp_guineapig?currentPage=all

Hopfer, S., and MacEachren, A. M. (2007). Leveraging the potential of geospatial annotations for collaboration: a communication theory perspective. *International Journal of Geographical Information Science*, *21*(8), 921-934.

ITO World. (2010, February 12). ITO World at TED 2010 - Project Haiti. *ITO world blog*. Retrieved September 18, 2010, from http://itoworld.blogspot.com/2010/02/ito-world-at-ted-2010-project-haiti.html

Jankowski, P., and Nyerges, T. (2001). GIS-Supported Collaborative Decision Making: Results of an Experiment. *Annals of the Association of American Geographers*, *91*(1), 48-70. doi:10.1111/0004-5608.00233

Jones, C. E. (2010). Practical cartography. In *Interacting with Geospatial Technologies*. Wiley-Blackwell.

Keim, D. A., Panse, C., and Sips, M. (2005). Information Visualization: Scope, Techniques, and Opportunities for Geovisualization. In *Exploring Geovisualization*. Pergamon.

Kidd, K. (2009, March 8). The Star unveils unique map of neighbourhoods. *The Star*. Retrieved from http://www.thestar.com/article/598502

Kingston, R. (2002). Web-based PPGIS in the United Kingdom. In *Community Participation and Geographical Information Systems* (1st ed.). CRC.

Kirakowski, J., and Corbett, M. (1990). *Effective Methodology for the Study of HCI*. Human factors in information technology. Amsterdam: North-Holland.

Klinkenberg, B. (2009). Citizen science and volunteered geographic information: can these help in biodiversity study? *Biodiversity of British Columbia*. Retrieved August 19, 2009, from http://www.geog.ubc.ca/biodiversity/VGI--VolunteerGeographicInformation.html

Kramers, E. R. (2008). Interaction with Maps on the Internet A User Centred Design Approach for The Atlas of Canada. *Cartographic Journal, The*, *45*, 98-107. doi:10.1179/174327708X305094

Kubota, Y. (2009, May 13). Google reshoots Japan views after privacy complaints. *Reuters*. Retrieved from http://www.reuters.com/article/idUSTRE54C22R20090513

Kyem, P. A. K. (2004). Of Intractable Conflicts and Participatory GIS Applications: The Search for Consensus amidst Competing Claims and Institutional Demands. *Annals of the Association of American Geographers*, *94*(1), 37-57. doi:10.1111/j.1467-8306.2004.09401003.x

Lake, R. W. (1993). Planning and applied geography: positivism, ethics, and geographic information systems. *Progress in Human Geography*, *17*(3), 404-413. doi:10.1177/030913259301700309

Lamentia, J. (2006). Second Generation Tag Clouds. Retrieved June 6, 2010, from http://www.joelamantia.com/ideas/second-generation-tag-clouds

Lamentia, J. (2007a). 10 Best Practices For Displaying Tag Clouds. *Joe Lamantia.com*. Retrieved July 28, 2010, from http://www.joelamantia.com/tag-clouds/10-best-practices-for-displaying-tag-clouds

Lamentia, J. (2007b). The Tagging Growth Curve. *You're It!* ». Retrieved August 5, 2010, from http://tagsonomy.com/index.php/the-tagging-growth-curve/

Lohmann, S., Ziegler, J., and Tetzlaff, L. (2009). Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I* (pp. 392-404). Uppsala, Sweden: Springer-Verlag. Retrieved from http://portal.acm.org/citation.cfm?id=1615856

MacEachren, A. M. (1994). Visualization in Modern Cartography: Setting the Agenda. In *Visualization in Modern Cartography*. Pergamon.

MacEachren, A. M. (2004). *How Maps Work: Representation, Visualization, and Design* (1st ed.). The Guilford Press.

MacEachren, A. M., Brewer, I., and Steiner, E. (2001). Geovisualization to mediate collaborative work: tools to support different-place knowledge construction and decision-making. *PROCEEDINGS, 20TH INTERNATIONAL CARTOGRAPHIC CONFERENCE*, 6-10.

MacEachren, A. M., and the ICA Commission on Visualization. (1998). VISUALIZATION — Cartography for the 21st century. In *Proceedings of the 7th Annual Conference of Polish Spatial Information Association*. Retrieved from http://www.geovista.psu.edu/sites/icavis/icavis/poland1.html

Maguire, D. (2007). GeoWeb 2.0 and Volunteered GI. Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/participants.html

Mahemoff, M. (2006). *Ajax Design Patterns* (1st ed.). Sebastopol, CA: O'Reilly.

Malherbe, A. (2009). Les données géographiques ont la cote. *Géomatique Expert*, *66*, 10-11.

Maron, M. (2010, January 14). Haiti OpenStreetMap Response. *OpenGeoData*. Retrieved September 18, 2010, from http://old.opengeodata.org/2010/01/14/haiti-openstreetmap-response/

Mathes, A. (2004, December). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Retrieved July 12, 2009, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

Maué, P. (2007). Reputation as tool to ensure validity of VGI. Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/participants.html

McCleary, G. F. (1975). In pursuit of the map user. In *Proceedings, Auto-Carto II* (pp. 238-250). Presented at the Auto-Carto, Washington, DC, US. Retrieved from http://mapcontext.com/autocarto/proceedings/auto-carto-2/pdf/in-pursuit-of-the-map-user.pdf

Meier, P. P. (2009, April 10). Developing Swift River to Validate Crowdsourcing « iRevolution. *iRevolution*. Retrieved December 8, 2009, from http://irevolution.wordpress.com/2009/04/10/developing-swift-river-to-validate-crowdsourcing/

Meng, Y., and Malczewski, J. (2009). Usability evaluation for a web-based public participatory GIS: A case study in Canmore, Alberta. *Cybergeo*. Retrieved from http://cybergeo.revues.org/index22849.html?

Michaud, N. (2008, September 22). OpenStreetMap : L'information géographique issue de la collaboration de masse. *Baliz-Media*. Retrieved August 15, 2009, from http://media.baliz-geospatial.com/fr/article/openstreetmap-l-information-geographique-issue-de-la-collaboration-de-masse

Moreno-Sanchez, R., Anderson, G., Cruz, J., and Hayden, M. (2007). The potential for the use of Open Source Software and Open Specifications in creating Web-based cross-border health spatial information systems. *INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE*, *21*(2007), 1135-1163.

Moullet, C. (2009, April 20). Future of MapFish Client and powerful queries. *MapFish*. Retrieved July 20, 2009, from http://mapfishblog.blogspot.com/2009/04/future-of-mapfish-client-and-powerful.html

Mummidi, L., and Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal*, *72*(3), 215-227. doi:10.1007/s10708-008-9181-5

Nichols, D. M., and Twidale, M. B. (2003). The usability of Open Source Software. *First Monday*, *8*(1).

Nielsen, J. (1993). *Usability Engineering*. Boston ; San Diego: Academic Press.

Nielsen, J. (1994). Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. *useit.com: usable information technology*. Retrieved October 5, 2010, from http://www.useit.com/papers/guerrilla_hci.html

Nielsen, J. (2005). 10 Heuristics for User Interface Design. *useit.com: usable information technology*. Retrieved June 6, 2010, from http://www.useit.com/papers/heuristic/heuristic_list.html

Nivala, A. M., Brewster, S., and Sarjakoski, L. T. (2008). Usability evaluation of Web mapping sites. *CARTOGRAPHIC JOURNAL*, *45*(2008), 129-138.

Noble, B. (2007). *Assessment of Affordable Housing Options using Collaborative Geospatial Software* (Master of Environmental Studies). University of Waterloo.

Noyle, B., and Bouwman, D. (2009). Usability and the GeoWeb: Part 1 of ? *GIS and .NET Development*. Retrieved June 24, 2010, from http://briannoyle.wordpress.com/2009/07/03/useability-and-the-geoweb-part-1-of/

Obermeyer, N. (2007). Thoughts on Volunteered (Geo)Slavery. Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/participants.html

Openshaw, S. (1991). A view on the GIS crisis in geography, or, using GIS to put Humpty-Dumpty back together again. *Environment & Planning A*, *23*(5), 621(8).

O'Reilly, T. (2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Retrieved January 8, 2009, from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1

Ottens, S. M. (2008, July 28). MapBuilder - End Of Life. Retrieved July 20, 2009, from http://communitymapbuilder.org/display/MAP/EndOfLife

Peng, Z., and Tsou, M. (2003). *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Network*. Wiley.

Perkins, C., and Dodge, M. (2008). The Potential of User-Generated Cartography: A Case Study of the OpenStreetMap Project and Mapchester Mapping Party. *North West Geography*, *8*(1), 19-32.

Pickles, J. (1991). Geography GIS and the surveillant society. *Papers & Proceedings of Applied Geography Conferences - State University of New York at Binghamton*, *14*, 80-91.

Pickles, J. (1995). *Ground Truth: The Social Implications of Geographic Information Systems* (1st ed.). The Guilford Press.

Pickles, J. (1999). Arguments, debates, and dialogues: the GIS-social theory debate and the concern for alternatives. In *Geographical Information Systems: Principles, Techniques, Applications and Management, 2nd Edition* (2nd ed.). Wiley.

Preece, J. (1993). *A Guide to Usability: Human Factors in Computing*. Addison Wesley.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). *Human Computer Interaction*. Addison Wesley.

Purves, R., Dykes, J., Edwardes, A., Hollenstein, L., Mueller, D., and Wood, J. (2009). Describing the space and place of digital cities through volunteered geographic information. Presented at the Geoviz, Hamburg.

Purves, R., and Edwardes, A. (2008). Exploiting Volunteered Geographic Information to describe Place. In *Proceedings of the GIS Research UK 16th Annual Conference GISRUK 2008*. Presented at the GISRUK, Manchester.

Quadro, F. (2007, July 19). MapServer versus GeoServer. Retrieved July 20, 2009, from http://www.fernandoquadro.com.br/html/2007/07/19/mapserver-versus-geoserver-english-version/

Ramsey, P. (2007). The state of Open Source GIS. Retrieved from http://www.refractions.net/expertise/whitepapers/opensourcesurvey/survey-open-source-2007-12.pdf

Ramsey, P. (2009, June 16). PostGIS versus MySQL Spatial at GeoSpiel. Retrieved July 17, 2009, from http://docs.opengeo.org/geospiel/2009/06/16/postgis-versus-mysql-spatial/

Raymond. (2000). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (1st ed.). O'Reilly Media, Inc.

Rinner, C., Keßler, C., and Andrulis, S. (2008). The use of Web 2.0 concepts to support deliberation in spatial decision-making. *Computers, Environment and Urban Systems*, *32*(5), 386-395. doi:10.1016/j.compenvurbsys.2008.08.004

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., and Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 995-998). San Jose, California, USA: ACM. doi:10.1145/1240624.1240775

Roberts, J. C. (2005). Exploratory Visualization with Multiple Linked Views. In *Exploring Geovisualization*. Pergamon.

Roberts, J. C. (2008). Coordinated Multiple Views for Exploratory Geovisualization. In *Geographic Visualization: Concepts, Tools and Applications*. Wiley Blackwell.

Robinson, A. (2009, September 20). Coordinated Views, Brushing, & Highlighting. *Cartography 2.0*. Retrieved December 24, 2009, from http://cartography2.org/Chapters/page14/CoordinatedViews.html

Robinson, A. C. (2006). Highlighting Techniques to Support Geovisualization. In *Proceedings of the ICA Workshop on Geovisualization and Visual Analytics*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.79.757

Roth, R. E. (2009, September 21). Cartographic Interaction Overview. *Cartography 2.0*. Retrieved December 24, 2009, from http://cartography2.org/Chapters/page13/MapInteractionOverview.html

Roth, R. E., and Ross, K. S. (2009). Extending the Google Maps API for Flash for Event Animation Mashups. *Cartographic Perspectives*, *64*, 21-40.

Rowe, G., and Frewer, L. J. (2005). A Typology of Public Engagement Mechanisms. *Science Technology Human Values*, *30*(2), 251-290. doi:10.1177/0162243904271724

Scharl, A. (2007). Towards the Geospatial Web: Media Platforms for Managing Geotagged Knowledge Repositories. In *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society* (1st ed.). Springer.

Schaub, T. (2009, July 20). OpenLayers Down Under. Retrieved July 20, 2009, from http://openlayers.org/blog/2009/07/20/openlayers-down-under/

Schlossberg, M., and Shuford, E. (2005). Delineating “Public” and “Participation” in PPGIS.

URISA, *16*(2), 15-26.

Seeger, C. (2008). The role of facilitated volunteered geographic information in the landscape planning and site design process. *GeoJournal*, *72*(3), 199-213. doi:10.1007/s10708-008-9184-2

Seifert, C., Kump, B., Kienreich, W., Granitzer, G., and Granitzer, M. (2008). On the beauty and usability of tag clouds. *PROCEEDINGS OF THE 12TH INTERNATIONAL INFORMATION VISUALISATION*, 17-25.

Sheesley, B. (2009). Data Probing and Info Window Design. *Cartography 2.0*. Retrieved February 3, 2010, from http://cartography2.org/Chapters/page8/DataProbing.html

Sheskin, I. (1985). *Survey research for geographers*. Association of American Geographers.

Sidlar, C. L., and Rinner, C. (2007). Analyzing the Usability of an Argumentation Map as a Participatory Spatial Decision Support Tool. *URISA Journal*, *19*(1), 47-55.

Sieber, R. (2006). Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers*, *96*(3), 507, 491.

Sieber, R. (2007). GeoWeb for Social Change. Presented at the Workshop on Volunteered Geographic Information, Santa Barbara, CA. Retrieved from http://www.ncgia.ucsb.edu/projects/vgi/supp.html

Sinclair, J., and Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful? *JOURNAL OF INFORMATION SCIENCE*, *34*(1), 15-29. doi:10.1177/0165551506078083

Skarlatidou, A., and Haklay, M. (2006). Public Web Mapping: Preliminary Usability Evaluation. In *Geographical Information Science Research Conference 2006 (GISRUK'06)*. Nottingham, UK. Retrieved from http://eprints.ucl.ac.uk/16228/

Slingsby, A., Dykes, J., Wood, J., and Clarke, K. (2007). Interactive Tag Maps and Tag Clouds for the Multiscale Exploration of Large Spatio-temporal Datasets. In *Proceedings of the 11th International Conference Information Visualization* (pp. 497-504). IEEE Computer Society. Retrieved from http://portal.acm.org/citation.cfm?id=1271665

Slocum, T. A. (1983). Predicting Visual Clusters on Graduated Circle Maps. *Cartography and Geographic Information Science*, *10*, 59-72. doi:10.1559/152304083783948168

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S., and Hedley, N. R. (2001). Cognitive and Usability Issues in Geovisualization. *Cartography and Geographic Information Science*, *28*, 61-75. doi:10.1559/152304001782173998

Slocum, T. A., McMaster, R. B., Kessler, F. C., and Howard, H. H. (2008a). *Thematic Cartography and Geovisualization (3rd Edition)* (3rd ed.). Prentice Hall.

Slocum, T. A., McMaster, R. B., Kessler, F. C., and Howard, H. H. (2008b). Proportionnal Symbol Mapping. In *Thematic Cartography and Geovisualization (3rd Edition)* (3rd ed., pp. 302-327). Prentice Hall.

Slocum, T. A., McMaster, R. B., Kessler, F. C., and Howard, H. H. (2008c). Map Animation. In *Thematic Cartography and Geovisualization (3rd Edition)* (3rd ed., pp. 389-408). Prentice Hall.

Smith, G. (2008). *Tagging: People-powered Metadata for the Social Web* (1st ed.). New Riders Press.

Straup Cope, A. (2008, October 30). The Shape of Alpha. *Code: Flickr Developer Blog*. Retrieved August 18, 2009, from http://code.flickr.com/blog/2008/10/30/the-shape-of-alpha/

Straup Cope, A. (2009, February 9). Things I'm Standing Next To. *Code: Flickr Developer Blog*. Retrieved August 18, 2009, from http://code.flickr.com/blog/2009/02/09/things-im-standing-next-to/

Sui, D. (2008). The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems*, *32*(1), 5, 1.

Tapscott, D., and Williams, A. D. (2006). *Wikinomics: How Mass Collaboration Changes Everything*. New York: Portfolio.

Taranu, J. (2009). *Building Consensus using a Collaborative Spatial Multi-Criteria Analysis System* (Master of Environmental Studies). University of Waterloo.

Taylor, P. J. (1990). Editorial comment GKS. *Political Geography Quarterly*, *9*(3), 212.

Tesquet, O. (2009, June 15). En Iran, "révolution Twitter" ou révolution tweetée? *L'express*. Retrieved from http://www.lexpress.fr/actualite/monde/proche-orient/en-iran-revolution-twitter-ou-revolution-tweetee_767535.html

The Economist. (2010, February 4). Data and transparency: Of governments and geeks | The Economist. Retrieved from http://www.economist.com/node/15469415?story_id=15469415

Tobón, C. (2005). Evaluating Geographic Visualization Tools and Methods: An Approach and Experiment Based upon User Tasks. In *Exploring Geovisualization*. Pergamon.

Tulloch, D. (2007, February 5). Many, many maps: Empowerment and online participatory mapping. *First Monday*. Retrieved August 9, 2009, from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1620/1535

Tulloch, D. (2008). Is VGI participation? From vernal pools to video games. *GeoJournal, 72*(3), 161-171. doi:10.1007/s10708-008-9185-1

Turner, A. (2006). *Introduction to Neogeography*. Short Cut (O'Reilly.).

Tuupola, M. (2008, November 4). Introduction to Marker Clustering With Google Maps. Retrieved July 21, 2009, from http://www.appelsiini.net/2008/11/introduction-to-marker-clustering-with-google-maps

Von Haaren, C., and Waren-Kretzmar, B. (2006). The interactive landscape plan: Use and benefits of new technologies in landscape planning and discussion of the interactive landscape plan in Koenigslutter am Elm, Germany. *Landscape Research, 31*(1), 83-105.

Wang, H., and Wang, C. (2001). Open Source Software Adoption: A Status Report. *IEEE Softw.*, *18*(2), 90-95.

Warren, D. (1991). *Using indigenous knowledge in agricultural development*. World Bank. Retrieved from http://ideas.repec.org/p/fth/wobadi/127.html

Weaver, C. (2004). Building Highly-Coordinated Visualizations in Improvise. *IEEE Symposium on Information Visualization, 2004*, 159-166.

Weaver, C. (2005). Visualizing coordination in situ. *Proceedings - IEEE Symposium on Information Visualization, 2005*.

Weiner, D., and Harris, T. M. (2003). Community-integrated GIS for Land Reform in South Africa. *URISA Journal*, *15*(2), 73, 61.

Weiner, D., and Harris, T. M. (2007). Participatory Geographic Information Systems. In *The Handbook of Geographic Information Science*. Wiley-Blackwell.

Wheatley, M. (2004). Open Source: The Myths of Open Source - CIO.com - Business Technology Leadership. Retrieved December 8, 2008, from http://www.cio.com/article/32146/Open_Source_The_Myths_of_Open_Source

Wheeler, D. A. (2007). Why Open Source Software / Free Software (OSS/FS, FOSS, or FLOSS)? Look at the Numbers! Retrieved December 8, 2008, from http://www.dwheeler.com/oss_fs_why.html

White, J. D., and Roth, R. E. (2010). TwitterHitter: Geovisual analytics for harvesting insight from volunteered geographic information. In *Proceedings of GIScience 2010*. Presented at the GIScience, Zurich, Switzerland.

White, P. (2008, October 3). Map-making mania. *Globe and Mail*. Retrieved from http://www.theglobeandmail.com/servlet/story/RTGAM.20081003.wlmapmakers03/BNStory/lifeMain/home

Zhang, J., and Goodchild, M. F. (2002). *Uncertainty in Geographical Information* (1st ed.). CRC.

Appendix A
Workshop material

University of Waterloo

May 1st, 2009

Dear

This letter is an invitation to attend a workshop I am conducting as part of my Master's degree in the Department of Geography at the University of Waterloo under the supervision of Professor Robert Feick.

Through a series of workshops involving many residents of the Bulkley Valley, a great deal of community input has been recorded in an online mapping tool called MapChat. One of the key features of MapChat is that it allows residents to share a common online map on which each person can express their concerns, knowledge and ideas by drawing points, lines and areas on the map and linking text comments to their drawings. To date, MapChat workshops have been held on topics such as riparian zones at risk, community trails, the Town of Smithers OCP and community asset identification.

We are interested in developing this software further to provide better capabilities for presenting and browsing the residents' mapped comments. Therefore, this new workshop aims to build upon these earlier workshops by assisting participants to identify features, areas, and topics of particular interests amongst the data previously collected. In particular, the study intends to help us investigate the value of the new tool called MapChat Visualization designed to facilitate the exploration of the contributed data.

Participation in this study is voluntary. It will involve an interactive computer-based workshop of approximately an hour to take place at a computer lab or another location of your convenience. The workshop will be held during the last two weeks of June. It consists in performing a few tasks along with answering a brief questionnaire. You may decline to participate in any of the tasks or answer any of the survey questions if you so wish. Further, you may decide to withdraw from this study at any time without any negative consequences by advising the researcher.

To enable us to understand better how well the software functions, some aspects of your use of the tool are recorded (the amount that you zoom in on the map, which tools you make use of, etc.). Note that no personal information is recorded in this process and that all information you provide is considered completely confidential. Your name will not appear in any thesis or report resulting from this study, however, with your permission anonymous quotations and anonymous results related to stored user inputs and user interactions with the software may be used. All written information collected from this session will not include personal identifying information and will be kept for a period of one year in a locked location at the University of Waterloo and digital information will be retained indefinitely on a secure password protected server at the University of Waterloo. Only researchers associated with this project will have access to this information. There are no known or anticipated risks to you as a participant in this study.

If you have any questions regarding this study, or would like additional information to assist you in reaching a decision about participation, please contact me by email at vdeparda@uwaterloo.ca. You can also contact my supervisor, Professor Robert Feick at (519) 888-4567 ext. 35493 or email rdfeick@uwaterloo.ca.

I would like to assure you that this study has been reviewed and received ethics clearance through the Office of Research Ethics at the University of Waterloo. However, the final decision about participation is yours. If you have any comments or concerns resulting from your participation in this study, please contact Dr. Susan Sykes of this office at (519) 888-4567 Ext. 36005 or ssykes@uwaterloo.ca.

I hope that the results of this study will benefit the organizations directly involved in the study, as well as the greater community. As researchers we hope to learn more about the process of meaning extraction from local volunteered knowledge and understand better how participants use the tools for further enhancements.

If you are interested in participating to this study, please contact me by email at vdeparda@uwaterloo.ca.

I very much look forward to seeing you at the workshop and thank you in advance for your assistance.

Yours Sincerely,

Vivien Deparday

University of Waterloo

Introduction

- Through a series of workshops involving residents of the Bulkley Valley, British Columbia, community input on several issues has been recorded in an online mapping tool called MapChat. One of the key features of MapChat is that it allows residents to share a common online map on which each person can record their knowledge of the local area by drawing points, lines and areas on the map and linking text comments to their drawings. To date, MapChat workshops have been held on topics such as riparian zones at risk, community trails, the Town of Smithers Official Community Plan and community asset identification.
- The purpose of this workshop and survey is to learn which methods of presenting residents' mapped knowledge and comments are easiest to use and which methods provide you with the most meaningful information in terms of the data and their presentation. Data from several of the Bulkley Valley workshops have been combined together for this exercise and integrated in MapChat Visualization, software developed for this workshop. Any references to individuals' names or other identifying text have been removed to preserve anonymity of the community members.
- The following scenario is used to conduct this workshop. You will assume the role of a community leader who is going to attend a meeting with planners and town managers from the Town of Smithers in the Bulkley Valley, British Columbia. You want to present some amenities of the Bulkley Valley that are of particular significance to the community. This workshop guides you through a sequence of tasks that will help you make good use of the data contributed by the people of the Bulkley Valley using the MapChat Visualization tool.
- First, you have to fill in and sign the form of consent attached to this document and then you can start by answering the few questions below.

About You:

1) Gender:

☐ Male ☐ Female

2) Age Group:

☐ 0-18 ☐ 18-24 ☐ 25-34 ☐ 35-44 ☐ 45-54 ☐ 55-64 ☐ 65+

3) Are you familiar with the Bulkley Valley area?

☐ Yes ☐ No

4) Did you participate in any of the Bulkley Valley workshops?

☐ Yes ☐ No

If yes, please indicate which workshop(s) you participated in

☐ Riparian ☐ Trails ☐ Asset identification ☐ Smithers OCP

5) How often do you use computer software and the Internet in a month?

☐ Never ☐ Sometimes (1 to 5 times) ☐ Regularly (5 to 15) ☐ Frequently (15 to 30) ☐ Very frequently

6) How often do you use computer-based maps in a month (e.g. Google Maps, MapQuest, GIS Software)?

☐ Never ☐ Sometimes (1 to 5 times) ☐ Regularly (5 to 15) ☐ Frequently (15 to 30) ☐ Very frequently

7) Have you ever been involved in local planning or land management issues in the Bulkley Valley or else where?

☐ Yes ☐ No

8) If yes, please explain in general terms the context of your involvement:

____

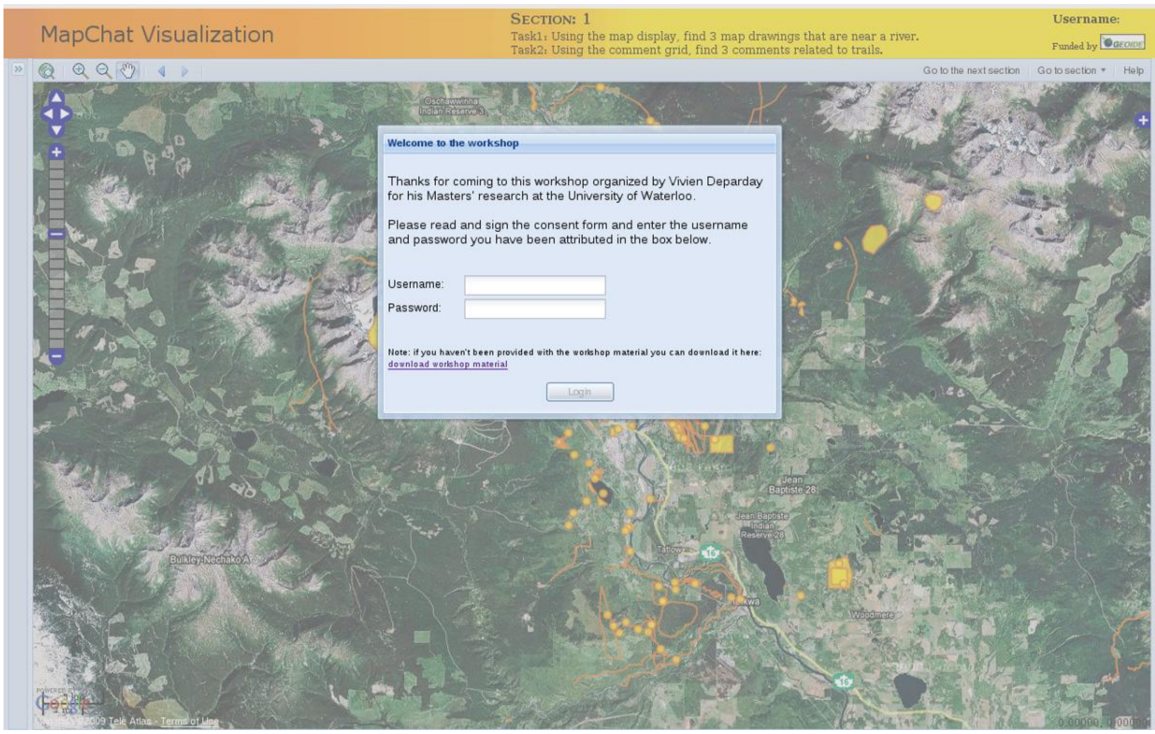
____

____

Section 1: Browsing the Data

This section of the workshop is designed to get you familiarised with the basic interface of MapChat Visualization (MapChat Viz) to navigate the map and the comments.

1. Go to the following address: http://athabasca.uwaterloo.ca/mapchatviz/
   You are presented with the following screen:



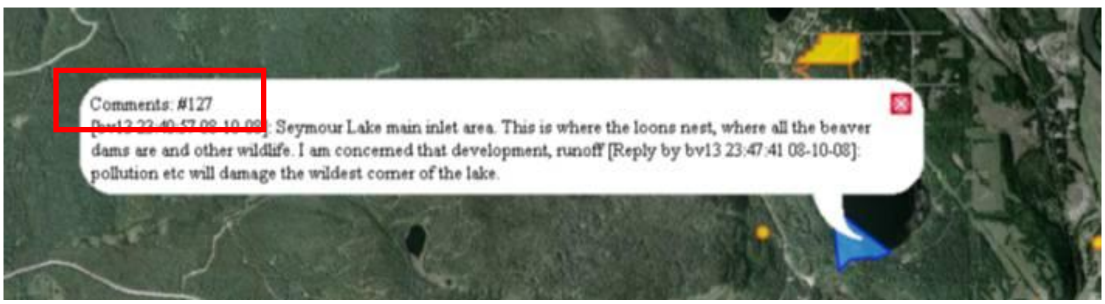
1. Enter the username you have been given at the top of the first page and the password given by the workshop facilitator in the two text boxes and click on the Login button.
2. You may want to press the F11 key on your keyboard to maximize the size of your web browser. You can come back to the initial view at any moment by simply pressing F11 again.
3. Before beginning the workshop tasks, you may want to experiment with the various map zoom tools on the tool bar (e.g. Zoom in, Zoom out, etc.).

   A description of these tools is provided in the first section of the MapChat Viz Help file.
4. Drawings of the locations of features contributed by participants at one or more of the Bulkley Valley MapChat workshops are displayed as orange points, lines or areas on the map. Note that in some cases throughout this workshop, one or more drawings will change colour to indicate that you have gone over them or that you have selected them.

5. To display a pop-up for a drawn feature, you need to click on the drawing when it is shaded in a light blue colour, as shown on the screen capture below.

Once you click on the drawing, a pop-up will appear containing the comment ID on the first line (framed in red on the image below) and the comment content.



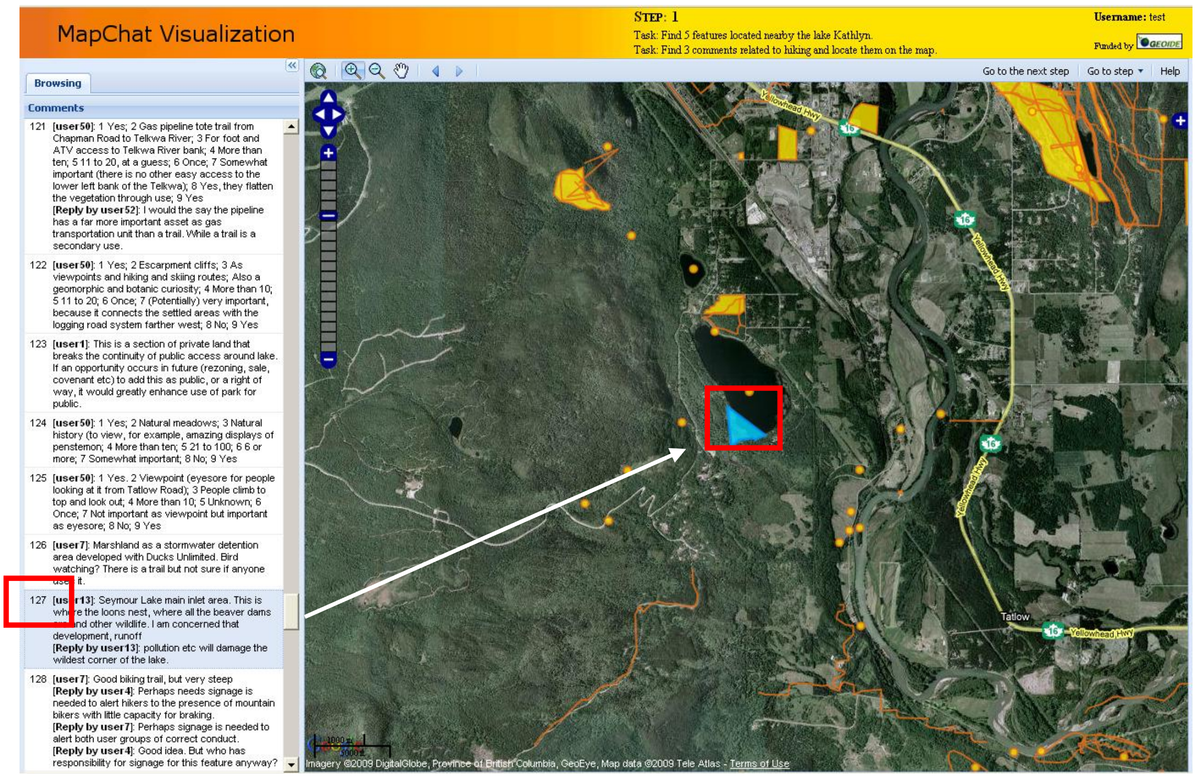
6. Experiment with this by opening a comment linked to a drawing by using a pop-up.
7. Now open the left column of the MapChat Viz interface by clicking on the double arrow on the top left-hand corner.

8. Once you do this you will see a tab that contains comments on the left of your screen. This tab contains by default all of the comments that all participants in the initial workshops made on the drawn features, as well as replies by themselves or other participants marked by a bold "reply" mark. In the left part of the tab, the id of the

comment (see screen capture below) is shown. It matches the one include in the linked drawing popup that you experimented with in Step 7.

9. Browse the comments situated in the comments tab. You will notice that when the mouse cursor moves over a comment in the tab, the associated drawn feature is highlighted in light blue on the map, as shown on the screen capture.

Comment id



10. Choose a comment and click your left mouse button on top of it. The map will move and zoom to the drawing linked to the comment you have just clicked on. If you want to go back to the previous map view you can click on the "Previous view button": ◀ on the right of the toolbar.

11. Once you have done this, please answer the questions below:

Questions:

1) **How easy is it to display a comment linked to a drawing by using the comment pop up tool?**

☐ Very Hard ☐ Hard ☐ Neutral ☐ Easy ☐ Very easy

2) **How easy is it to locate a drawing on the map related to a given comment by using the comment tab?**

☐ Very Hard ☐ Hard ☐ Neutral ☐ Easy ☐ Very easy

Task 1: **Using the MapChat Viz map display, find 3 feature drawings (either points, lines or areas) that are near a river and record their ID numbers in the table below:**

#id:	#id:	#id:

Task 2: **Using the comment tab, find 3 comments related to trails and record their ID numbers in the table below:**

#id:	#id:	#id:

3) Please note any difficulties you encountered while browsing the map (task 1):

____

____

____

4) Please note any difficulties you encountered while browsing the Comment Tab (task 2):

____

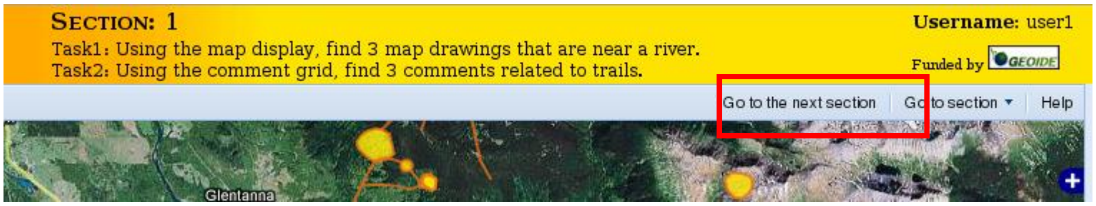
____

____

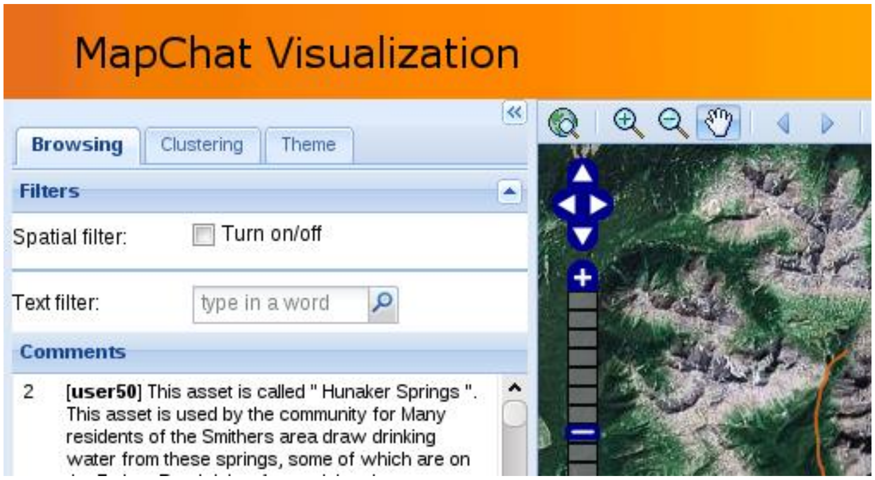
Section 2: Filtering the data

Finding relevant information among a large volume of mapped comments can be difficult. In this section, you will explore two features that attempt to improve the ease of data browsing.

1. Left click on the button **Go to the next section** on the top right hand corner of the screen.
   NOTE: At any moment during the workshop you can go back to the step you wish by using the "Go to step" drop down menu.



You will notice that a new panel called "Filters" has been opened in the Browsing tab on the left part of the screen. Two types of filters are available for you to use, a spatial filter and text filter as pointed out on the screenshot below. The spatial filter allows you to spatially filter the comments in the tab by showing only the comments that are linked to the drawings visible on the screen (so if you have zoomed to a smaller area, only these comments will be visible in the tab). The text filter allows you to filter the tab content according to a key word that you can type into the dialog box to the left of the magnifier. These features are further explained below and in the Step 2 of the MapChat Viz Help files.



2. To activate the Spatial filter, check the box labelled "Turn on/off" by clicking the left mouse button in the box. The content of the comment tab is now refreshed to contain only the comments linked to feature drawings visible on the screen. The comment tab content is refreshed every time the map view is changed.
3. Zoom in to an area of interest to you. Observe how the number of comments in the left-hand panel changes while you pan the map or zoom in or zoom out.

Task 1: **Using the spatial filter, find 3 map drawings that are near a lake and record the ID numbers of the 3 drawings in the table below:**

#id:	#id:	#id:

4. Turn the spatial filter off by left clicking again on the box. Feel free to use it later on in the workshop.
5. The text filter works by searching through the comments for the keyword you type into the search box. The comment tab will be refreshed to show only the comments that contain the keyword you entered. To use the text filter, type in a word and then left click on the magnifying glass or press enter. The tab content is refreshed and the drawings associated with the key word are highlighted in dark blue on the map.
   NOTE: You can then clear the search by clicking on the 'X' button beside the magnifying glass.
   NOTE: The spatial filter and text filter can be used at the same time. In this case, the comments shown in the grid will be filtered according to both filters at the same time.

Task 2: **Using the text filter, find 3 comments related to wildlife. Please record the ID numbers of the 3 comments related to wildlife that you selected:**

#id:	#id:	#id:

Please note the keyword(s) you searched for:

____

____

Questions:

1) **Does the spatial filter make browsing the comments attached to features drawn in a given area easier or harder than the manual method used in section 1 (Section 1 task 1 compared with Section 2 task 1)?**

☐ Much harder ☐ Harder ☐ Neutral ☐ Easier ☐ Much easier

2) **Does the text filter make finding the comments related to a key word and locating the associated features drawn on the map easier or harder than the manual method (Section1 task 2 compared with Section 2 task 2)?**

☐ Much harder ☐ Harder ☐ Neutral ☐ Easier ☐ Much easier

3) **Do you have any comments on the ease of using spatial filtering?**

____

____

____

____

____

4) **Do you have any comments on the ease of using text filtering?**

____

____

____

____

____

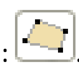
Section 3: Identifying areas of interest and summarising comments

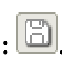
Determining areas with particularly high concentrations of user drawn objects with attached comments can help identify areas which are of particular importance to participants and which may necessitate a particular attention. Moreover, a map that contains many drawings can appear cluttered and difficult to understand. By zooming in, it is possible to display higher levels of detail for a smaller area. The purpose of this step is to look at different ways to represent concentrations of mapped comments.

<u>Section 3a: Manual method</u>

1. Left click on the button 

<u>Task 1:</u> **Use the drawing tools explained below to draw a bounding box around 3 areas that you think have an important concentration of comments.**

The drawing tools available in the MapChat Viz toolbar include four options:
. To define an area use the polygon tool: . To draw a polygon, click once on each point of the area you want to draw, and double-click on the last point. At each intermediate stage you will see the area as defined so far in an orange color. Once the polygon is finished, it appears in red.
Note: that the drawing tools to edit and delete your drawing are further explained in the third section of the Help file.

Don't forget to save your drawings by clicking on the save icon: **.**

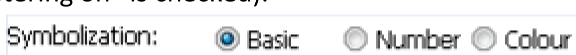
Section 3b: Assisted method:

It is not always easy to identify an area of concentration simply by looking at the mapped comments because some comments pertain to large areas or long lines, while others refer to a specific location that is symbolised with a point. One way of summarising the concentration of residents' comments on the map is now explained.

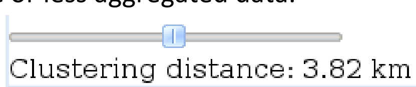
1. Again, left click the button 

2. Notice that the new tab called "clustering" appears in the left panel of the MapChat Viz interface. These functions are explored in the following step. They are also explained in the third part of the Help file.

3. Left click in the "Turn clustering on/off" checkbox. You will notice that some of the drawings are clustered together in green point symbols whose size depends on the number of features that are clustered in the point. The clustering process aggregates the drawings based on the clustering distance specified on the left (see point 5 below). If the drawings are within this distance threshold they are clustered in the same point. Also note that the clustering changes in relation to the map scale you are zoomed to. You can experiment with this by zooming in and out on the map. To see the full effect of what the clustering does, feel free to experiment by turning it off and then back on.
4. Try out the 3 different symbolization methods by clicking on each radio button: basic, number (the number of features contained in a displayed point) and colour. Note: make sure that "Turn clustering on" is checked).



5. Experiment with the distance slider to change the aggregation distance of the cluster. This way you can get more or less aggregated data.



6. Click on one of the cluster circles. You will notice that the circle turns blue as shown below and that a popup window appears. The cluster popup window is slightly different from the regular one. It allows you to browse through all the comments linked to the drawings clustered in the point by using the "next" and "prev" button at the bottom of the popup.

<u>Task 1:</u> **Now by using the clustering and the features available before, repeat the task 1 of Section 3a and circle 3 areas that you think have an important concentration of comments by using the drawing tools.**

Don't forget to save your drawings by clicking on the save icon: .

Questions:

1) Does the clustering method make the map more legible (i.e. easier to read) or less legible than the regular display (without clustering)?

☐ Much harder ☐ Harder ☐ Neutral ☐ Easier ☐ Much easier

2) Depending on your answer to the previous question, what aspect(s) of the visualization make(s) the use of clustering easier or more difficult?

____

____

____

3) Which symbols best represent the concentration of map drawings?

☐ Basic ☐ With numbers ☐ With colors and numbers

4) Does the clustering method make determining areas of concentration easier or harder to visualize and understand than the manual method used in Section3a?

☐ Much harder ☐ Harder ☐ Neutral ☐ Easier ☐ Much easier

5) Please explain your answer to question 4.

____

____

____

6) Do you believe that the clustering method depicts concentrations of users' drawings more accurately or less accurately than the manual method used in Section3a?

☐ Much less accurate ☐ Less accurate ☐ Neutral ☐ More accurate ☐ Much more accurate

7) Please explain your answer to question 6.

____

____

____

8) How complicated did you feel that the clustering feature was?

☐ Very complicated ☐ Complicated ☐ Neutral ☐ Simple ☐ Very simple

9) If you answered 'very complicated' or 'complicated' to the previous question, what might help to make this easier?

____

____

____

Section 4: Identifying and locating themes of interest

In this section, you will focus on identifying and locating themes and data categories in the information that is mapped. Hence, you will evaluate a way to identify themes and data categories that are identified as important by participants from the local community. To do this, one or more tags were added to each mapped feature in an attempt to characterise the different categories present in the comment(s) attached to the feature.

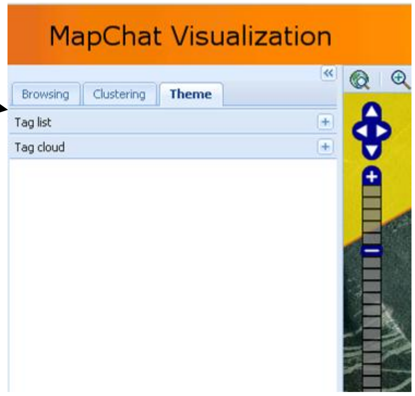
You can turn off the clustering by clicking on "Turn clustering on/off" to clear the checkbox

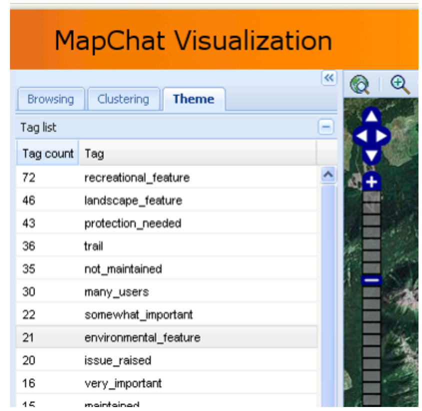
1. Left click the button 

2. Notice the new tab called "Theme" in the panel to the left of the map view. The functions in this tab are explored below. They are also explained in the fourth step of the Help file.
3. Display a popup for a feature by left clicking on a drawing within the map display. Notice the tags applied to this comment at the bottom of the pop-up.

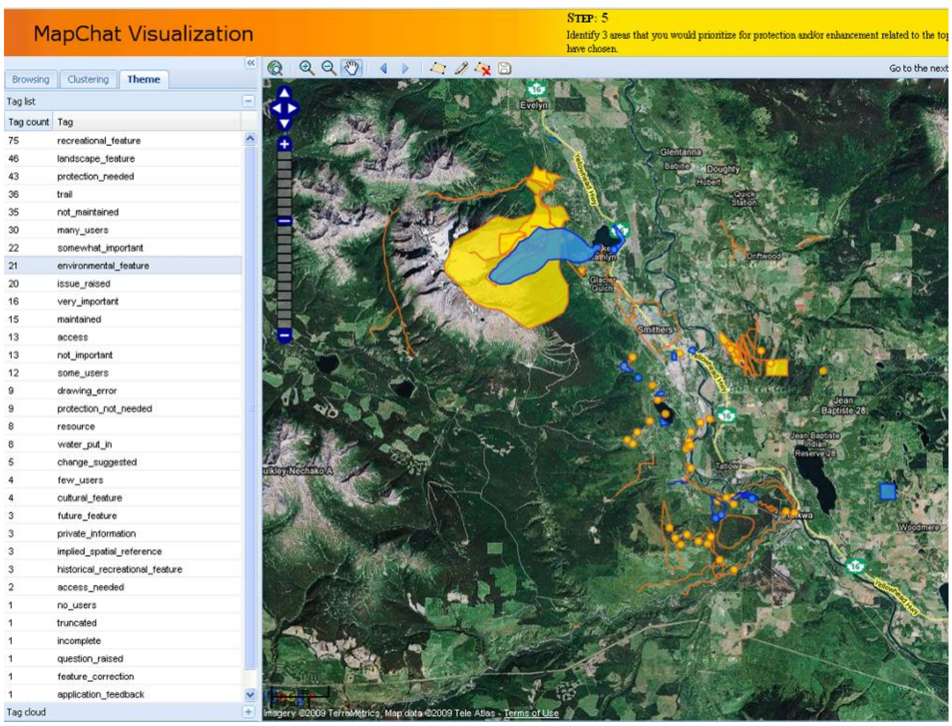
4. Left click on the tag list bar in the theme tab to open it and review the various tags ranked by their importance (the number of times each tag has been use to describe the different features in the map view).

Tag list bar

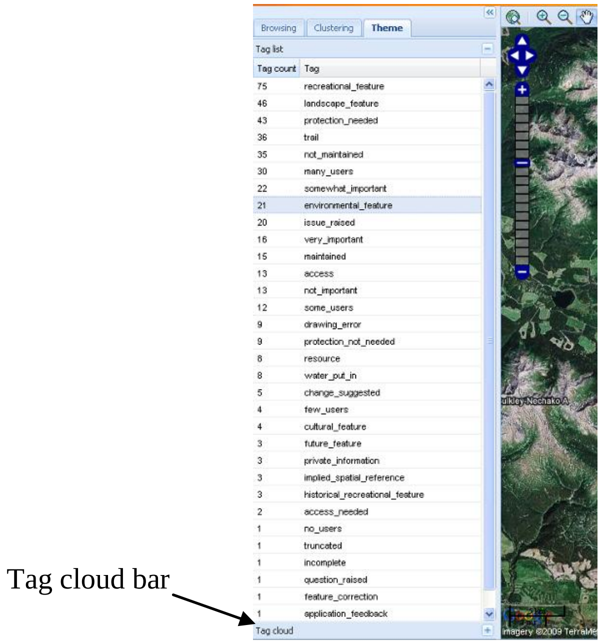


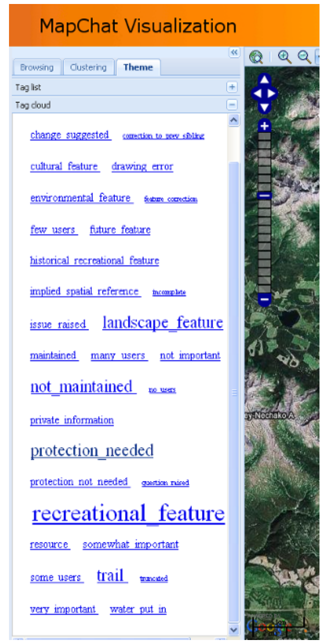


5. Hover the mouse cursor over a tag name and observe that all drawings related to this theme are highlighted on the map (see below).

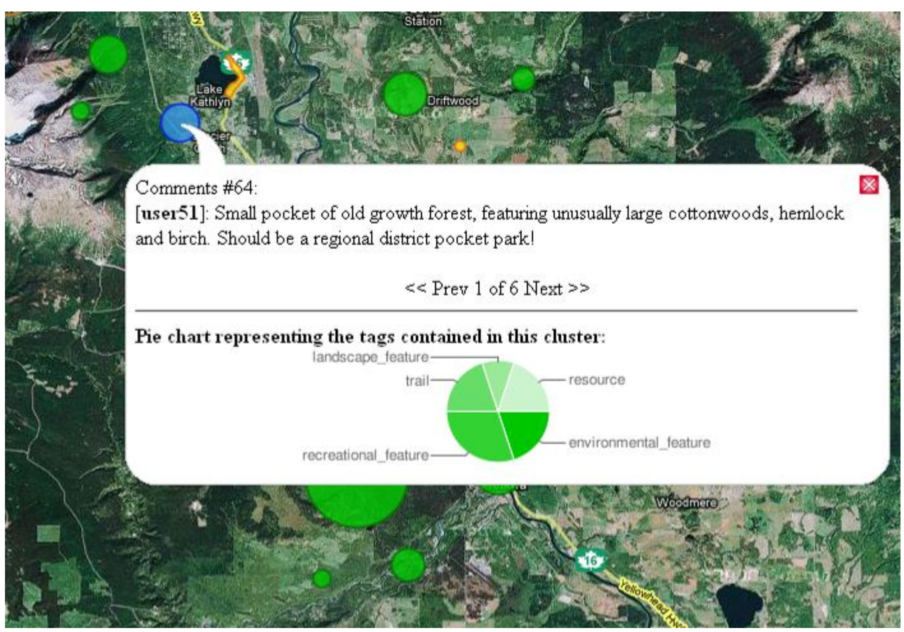


6. Zoom in, Zoom out and pan the map around and observe how the tag list content changes, as the count applies only to the features shown in the current map view. To see all of the tags set the map extent to full extent, with the far left icon on the tool bar.
7. Now left click on the cloud bar in the bottom right-hand corner of the bar to open it. The difference between the tag cloud and the tag list is that in the tag cloud the importance of a tag (tag count) is revealed by the size of the tag name. The tag names are ordered from top to bottom in alphabetical order.





8. Experiment with the tag cloud in the same way as the tag list. Hover the mouse cursor on a tag name and notice what features are highlighted on the map view; zoom in and out, pan around to see the tag cloud content changing.
9. Click on the clustering tab and turn the clustering back on. Click on one of the clusters to display a popup containing a pie chart. This pie chart represents the proportion of the tags that are applied to the comments contained in the chosen cluster.



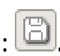
Task 1: **Choose an area you find interesting. Use the drawing tool as before to define this area by drawing a boundary around it. By using the tag cloud, tag list and the pie chart as you wish, review the themes for this area and note them below in order of importance (NOTE: a theme name can be tag name or any theme name of your choice):**

Theme1 : Theme 4:
Theme2 : Theme 5:
Theme3 : Theme 6:
Other themes or further comments:

____

____

____

Don't forget to save your drawings by clicking on the save icon: .

Questions:

1) How easy is it to review the themes of an area of interest in the above task (Section 4 task 1)?

☐ Very Hard ☐ Hard ☐ Neutral ☐ Easy ☐ Very easy

2) Please check the box that best reflect your impression of the tag list when achieving the task above (Section 4 task 1):

	Very much	Somewhat	Neutral	Somewhat	Very much	
Not useful	☐	☐	☐	☐	☐	Useful
Complicated	☐	☐	☐	☐	☐	Simple

3) Please check the box that best reflect your impression of the tag cloud when achieving the task above (Section 4 task 1):

	Very much	Somewhat	Neutral	Somewhat	Very much	
Not useful	☐	☐	☐	☐	☐	Useful
Complicated	☐	☐	☐	☐	☐	Simple

4) **Please check the box that best reflect your impression of the pie chart when achieving the task above (Section 4 task 1):**

	Very much	Somewhat	Neutral	Somewhat	Very much	
Not useful	☐	☐	☐	☐	☐	Useful
Complicated	☐	☐	☐	☐	☐	Simple

5) **Which one of the tag list and tag cloud in the previous section do you think best illustrate the relative importance of the different tags?**

☐ Tag list ☐ Tag cloud ☐ Neither

6) **Do you have any further comments regarding this task (Section 4 task 1)?**

____

____

____

____

____

____

Section 5: Applying what you have learned with MapChat Viz

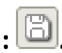
1. Left click the button **Go to the next section**

2. Choose one of the themes of interest listed below that you will use to complete this task:

☐ Wildlife conservation ☐ Outdoor recreation ☐ Other (Specify below)

Other: ____

<u>Task 1:</u> **By using any of the features that you have experienced through the different tasks in the workshop, identify a maximum of 3 areas related to the theme of interest you have chosen that you would prioritize for special attention in the Bulkley Valley. Use the drawing tool as before to identify the areas.**

Don't forget to save your drawings by clicking on the save icon: **.**

Questions:

1) Please rate how useful each tool was to achieve the task describe above (Section5 – task1) on a scale of 1 (not useful at all) to 5 (very useful):

	Not useful at all 1	2	Neutral 3	4	Very useful 5
Spatial filter	☐	☐	☐	☐	☐
Text filter	☐	☐	☐	☐	☐
Clustering	☐	☐	☐	☐	☐
Clustering distance	☐	☐	☐	☐	☐
Tag list	☐	☐	☐	☐	☐
Tag cloud	☐	☐	☐	☐	☐
Tag pie chart	☐	☐	☐	☐	☐
Pop up	☐	☐	☐	☐	☐
Zoom to a drawing by clicking on a comment	☐	☐	☐	☐	☐

2) Can you suggest any further means of visualizing and browsing the community data used in this exercise?

____

____

____

____

____

____

3) Do you have any further general comments on the workshop?

____

____

____

____

____

____

Thank you for participating in this study! Do not hesitate to request further information or to provide more feedback now or later by contacting me (vdeparda@uwaterloo.ca). Summarised results will be sent to you once they have been compiled and analysed.